Molecular Basis of Marfan Syndrome: *In vitro* and *in silico* Analyses of Exonic and Intronic Sequence Variants in the *FBN1* Gene

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)
vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

István Magyar

aus

Ungarn

Promotionskomitee

Prof. Dr. rer. nat. Wolfgang Berger (Vorsitz)

PD Dr. sc. nat. Gábor Mátyás (Leitung der Dissertation)

Prof. Dr. rer. nat. Thierry Hennet

Prof. Dr. sc. nat. Dieter Zimmermann

Zürich, 2011

Declaration

I declare that my thesis was composed by myself and the enclosed experimental works were performed on my own as indicated in the respective chapters. This dissertation has not been submitted for any other degree or professional qualification except as specified.

István Magyar, Zurich 2011

Abbreviations

A	Adenine
ACTA2	Actin, alpha 2, smooth muscle, aorta; OMIM 102620
ADAMTS10	A disintegrin and metalloproteinase with thrombospondin motifs 10; OMIM 608990
ADAMTSL4	A disintegrin and metalloproteinase with thrombospondin motifs-like 4; OMIM 610113
ATS	Arterial tortuosity syndrome; OMIM 208050
bp	Base pair
BTZ	Barentsz protein
C	Cytosine
Ca	Calcium
CCA	Congenital contractural arachnodactily
cDNA	Complementary DNA
CERES	Composite exonic regulatory element of splicing
COL2A1	Collagen, type II, alpha 1; OMIM 120140
COL3A1	Collagen, type III, alpha 1; OMIM 120180
COL9A1	Collagen, type IX, alpha 1; OMIM 120210
COL11A1	Collagen, type XI, alpha 1; OMIM 120280
COL11A2	Collagen, type XI, alpha 2; OMIM 120290
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
eIF4AIII	Eukaryotic translation initiation factor 4A, isoform 3
EJC	Exon junction complex
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
ER	Endoplasmatic reticulum
FBN1	Fibrillin 1; OMIM 134797
FBN2	Fibrillin 2; OMIM 121050
G	Guanine
gDNA	Genomic DNA
GLUT10	Protein encoded by *SLC2A10*
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
kDa	Kilo Dalton
LAP	Latency associated peptide
LDS	Loeys-Dietz syndrome; OMIM 609192
LLC	Large latent complex
LTBP	Latent transforming growth factor beta binding protein
LTBP1	Latent transforming growth factor beta binding protein 1; OMIM 150390
M	Molar
MAGOH	Mago-nashi homolog protein (Drosophila)
MFS	Marfan syndrome; OMIM 154700
MR	Mitral regurgitation
mRNA	Messenger RNA
mut	Mutant
MYH11	Myosin, heavy chain 11, smooth muscle; OMIM 160745

MVP	Mitral valve prolapse
NCBI	National Center for Biotechnology Information
NMD	Nonsense-mediated mRNA decay
OMIM	Online Mendelian Inheritance in Man
PBS	Phosphate buffered salt
PTC	Premature termination codon
RNA	Ribonucleic acid
RNPS1	RNA binding protein S1
RPGR	Retinitis pigmentosa GTPase regulator; OMIM 312610
RT-PCR	Reverse transcriptase polymerase chain reaction
SGS	Shprintzen-Goldberg craniosynostosis syndrome; OMIM 182212
SLC	Small latent complex
SLC2A10	Solute carrier family 2 (facilitated glucose transporter), member 10; OMIM 606145
SLC16A12	Solute carrier family 16 (monocarboxylic acid transporter), member 12; OMIM 611910
SMADs	MAD (mothers against decapentaplegic) homologues
SMG1	SMG1 homolog, phosphatidylinositol 3-kinase-related kinase (C. elegans)
SMN1	Survival of motor neuron 1
snRNP	Small nuclear ribonucleoprotein
SNuPE-ONCE	Single-nucleotide primer extension (SNuPE) and laser-induced fluorescence capillary electrophoresis (LIF-CE)
SURF	Surfeit locus protein
T	Thymine
TE	Tris-EDTA buffer
TAAD	Familial thoracic aortic aneurysms and dissections; OMIM 607086, 608967
TB	Tranforming growth factor beta-1 binding protein
TGFβ-1	Tranforming growth factor beta 1
TGFBR1	Transforming growth factor, beta receptor I; OMIM 190181
TGFBR2	Transforming growth factor, beta receptor II; OMIM 190182
U	Uracil
UPF1	UPF1 regulator of nonsense transcripts homolog (yeast)
UPF2	UPF2 regulator of nonsense transcripts homolog (yeast)
UPF3a	UPF3 regulator of nonsense transcripts homolog a (yeast)
UPF3b	UPF3 regulator of nonsense transcripts homolog b (yeast)
UTR	Untranslated region
Y14	RNA binding motif protein 8A
VDJ exon	Variable (V), Diversity (D), and Joining (J) exon of the immunoglobulin-μ gene
wt	Wild-type

Table of Contents

Summary

Heterozygous mutations in the human fibrillin-1 gene (*FBN1*, OMIM 134797) cause Marfan syndrome (MFS, OMIM 154700), an autosomal dominant connective tissue disorder, characterized by variable manifestations in the cardiovascular, ocular, and skeletal systems. For routine genetic testing, various mutation detection techniques have been developed using patients´ genomic DNA (gDNA) as template. The consequence of exonic variants and classical splice site defects can be predicted on the basis of the gDNA sequence. However, in routine diagnostic settings, the pathogenic consequences of mutations in regulatory elements, non-canonical splice donor and acceptor sites as well as deep intronic regions are more difficult to evaluate in the absence of RNA samples. Functional testing of all variants detected on gDNA level is difficult and unrealistic. In order to restrict the transcript analysis to those variants which possibly affect pre-mRNA splicing, the use of bioinformatic prediction-tools can be very helpful.

The central issue of this PhD thesis was to investigate potentially pathogenic consequences of a large number of exonic and intronic sequence variants in the *FBN1* gene on mRNA level. 93 novel and 32 previously known *FBN1* exonic mutations as well as intronic sequence variants were characterized by reverse transcription (RT)-PCR and subsequent fragment and/or sequence analyses. Furthermore, three online available exonic splicing enhancer (ESE) and five splice site prediction tools were applied and their predictions were compared with *in vitro* results. From this large collection of sequence alterations, splicing aberrations were detected for 27 variants including one silent mutation, two missense alterations, one small insertion, and 23 intronic substitutions. Bioinformatic approaches for ESE predictions proved to be not very reliable, because consequences of only one out of four aberrations were accurately predicted. For intronic substitutions within canonical splice site sequences, the online tools were found to be more reliable to assess potential aberrations (Paper 1; Magyar et al., 2011, manuscript in preparation).

In a further study, we improved, evaluated, and used Sanger sequencing for quantification of single nucleotide polymorphisms (SNPs) in mRNA transcripts and gDNA samples. Further, we used this procedure to analyse 26 out of 37 *FBN1* sequence variants leading to premature termination codons (PTCs). Comparative analysis of transcripts derived from primary dermal fibroblasts of affected patients with and without nonsense-mediated mRNA decay (NMD) inhibition as well as from fresh and PAXgene-stabilized blood samples demonstrated tissue-specific degradation of PTC-containing *FBN1* transcripts, providing evidence for incomplete NMD in leukocytes (Paper 2; Magyar et al., 2009).

In addition, we screened the gDNA of 101 unrelated individuals with suspected MFS, in whom standard genetic testing detected no mutation, for copy number variations by using multiplex ligation-dependent probe amplification (MLPA) and the Affymetrix Human Mapping 500K Array set. We successfully applied these methods to identify large *FBN1* deletions in two patients from our cohort. Subsequent sequence analysis of the deletion breakpoint fragments identified deletions in the putative regulatory and promoter region of *FBN1*, indicating that no mRNA will be transcribed from the deleted allele. This was confirmed by experimental transcript analyses (Paper 3; Mátyás et al., 2007).

The advantages and disadvantages of transcript analyses and the reliability of *in silico* prediction tools are discussed.

Zusammenfassung

Heterozygote Mutationen im humanen Fibrillin-1 Gen (*FBN1*, OMIM 134797) sind mit der autosomal-dominant vererbten Bindegewebserkrankung ´Marfan Syndrom´ (MFS, OMIM 154700) assoziiert. Das klinische Bild manifestiert sich durch unterschiedlich ausgeprägte Veränderungen im Herzen, im Blutgefässsystem, den Augen und im Skelett. Viele Methoden wurden für die routinmässige Detektion von Mutationen etabliert, allerdings beruhen alle auf der Untersuchung genomischer DNA (gDNA) der betroffenen Patientinnen und Patienten. Die Auswirkungen exonischer und klassischer Spleissstellen-Mutationen kann man, basierend auf der gDNA-Sequenz, vorhersagen. Im Gegensatz dazu ist im diagnostischen Routinetest das Vorhersagen der pathogenen Auswirkungen von Mutationen in regulatorischen Elementen, nicht kanonischen Akzeptor- und Donorspleisssstellen sowie in tiefen intronischen Bereichen des Gens, ohne zusätzliche RNA-Untersuchung, schwer zu bestimmen. Funktionelle Untersuchungen aller Mutationen sind schwierig und unrealistisch, da das entsprechende Patientenmaterial nicht immer vorliegt. Um sich auf diejenigen Proben für RNA-Untersuchung zu konzentrieren, bei denen eine Auswirkung auf mRNA-Ebene wahrscheinlich ist, könnten Vorhersageprogramme eine Hilfe bieten.

Im Rahmen dieser Dissertation wurden die pathogenen Auswirkungen einer grossen Anzahl von exonischen und intronischen *FBN1* Sequenzvarianten auf mRNA-Ebene charakterisiert. 93 neue und 32 bekannte exonische *FBN1* Mutationen sowie intronische Abweichungen wurden durch reverse Transkriptase (RT)-PCR und anschliessend mittels Fragment- und/oder Sequenz-Analysen untersucht. Drei exonische Spleissverstärker- (ESE) und fünf Spleisssstellen-Vorhersageprogramme wurden angewandt und die daraus resultierenden Vorhersagen mit *in vitro* erhaltenen Ergebnissen verglichen. Wir fanden für 27 Varianten Spleisseffekte, einschliesslich einer stillen Mutation, zwei missense Mutationen, einer kleinen Insertion und 23 intronischer Nukleotidaustausche. Dabei stellte sich heraus, dass bioinformatische Algorithmen für die Vorhersage von exonischen Spleissverstärkern nicht sehr zuverlässig sind, da nur eine der vier Spleissstörungen vorhergesagt wurde. Betreffend der intronischen Abweichungen in kanonischen Spleisssstellen waren die verwendeten Online-Programme zuverlässiger (Paper 1; Magyar et al., 2011, Manuskript in Vorbereitung).

In einer weiteren Studie wurde die Sanger-Sequenzierung von uns angewendet und mit dem Ziel optimiert, mögliche Effekte von Einzelnukleotid-Polymorphismen (SNPs) sowohl auf Transkript-, als auch auf gDNA-Ebene quantifizieren zu können. Diese Methode wurde anschliessend für 26 von 37 *FBN1* Sequenzvarianten mit vorzeitigen Stopkodons (PTC) angewandt. Eine vergleichende Analyse der aus Fibroblasten stammenden Transkripte von betroffenen Patientinnen und Patienten mit und ohne Inhibition des Nonsens-vermittelten mRNA Abbau-Prozesses (NMD) sowie aus frischen und PAXgene-stabilisierten Blutproben zeigten einen gewebespezifischen Abbau der *FBN1*-Transkripte. Diese Ergebnisse weisen auf einen unvollständigen NMD-Prozess in Leukozyten hin (Paper 2; Magyar et al., 2009).

In vielen Patienten mit Verdacht auf MFS konnte mit diagnostischen Routineuntersuchungen keine Mutation im *FBN1* Gen gefunden werden. Die genomische DNA von 101 dieser nicht-verwandten Patienten wurde auf mögliche Änderungen in der Kopienzahl von *FBN1*-Sequenzen untersucht. Dies erfolgte mittels Multiplex Ligation-Dependent Probe Amplifikation (MLPA) und Affymetrix Human Mapping 500K Array. Dabei konnten wir zwei grosse *FBN1* Deletionen identifizieren. Die Sequenzierungen der genauen Bruchstellen ergaben, dass die Deletionen in potentiell transkriptionsregulatorischen bzw. in den Promoterregionen lagen. Dies deutet darauf hin, dass keine mRNA vom deletierten Allel transkribiert wird, was durch experimentelle Transkriptanalysen bestätigt wurde (Paper 3; Mátyás et al., 2007).

Die Vorteile und Nachteile von Transkriptanalysen und die Zuverlässigkeit der verwendeten *in silico* Vorhersage-Programme werden diskutiert.

1 General Introduction

1.1 History of Marfan Syndrome

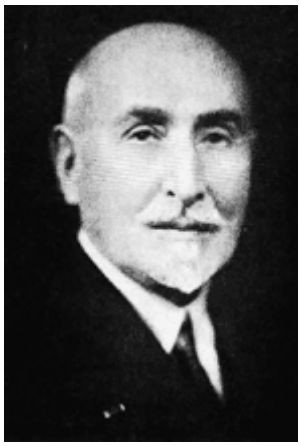

Figure 1. Antoine Bernard -Jean Marfan (1858-1942) (http://www.marfansresearchfoundation.ie/information.html).

In 1896, Antoine Bernard-Jean Marfan (Figure 1), a french pediatrician, described and presented the case of a little girl who was unusually tall and had exceptionally long limbs and asthenic physique [Marfan, 1896]. Her fingers and toes were disproportionaly long termed it spider-like (arachnodactyly). A few years later, another girl was described with similar abnormalities by Achard, also with long digits and articular hypermobility [Achard, 1902]. Recently, aware of the expanded phenotypic knowledge, it seems to be that Achard`s patient had Marfan syndrome and the little girl reported by Marfan suffered rather from congenital contractural arachnodactyly (CCA), which shows similar phenotypes without cardiovascular difficulties [Hecht and Beals, 1972].

The term Marfan syndrome or Marfan`s syndrome was introduced in 1931 by the Dutch ophthalmologists Henricus Jacobus Marie Weve. In 1972, Victor A. McKusick wrote a comprehensive review about the syndrome in his classic monograph Heritable Disorders of Connective Tissue. The breakthrough in the field of Marfan syndrome research followed after the identification of the fibrillin-1 protein from medium of human fibroblast cell cultures in the late 80s [Sakai et al., 1986]. A few years later, the first mutation was identified in the *FBN1* gene, which was mapped previously to chromosome 15q15-23.1 [Kainulainen et al., 1990; Dietz et al., 1991a; Dietz et al., 1991b]. The second Marfan locus (MFS2) was mapped to 3p24.2-p25 in a large french family and the gene was thought to play a role in the TGFβ (Transforming growth factor beta) signaling, which was confirmed by the identification of MFS2-causing *TGFBR2* (TGFβ receptor II) mutations in 2004 [Collod et al., 1994; Neptune et al., 2003; Mizuguchi et al., 2004]. Subsequently, the role of the *TGFBR1* (TGFβ receptor I) gene was shown in patients with suspected MFS who were negative for the *FBN1* and *TGFBR2* gene mutations [Loeys et al., 2005; Mátyás et al., 2006; Singh et al., 2006b]. In fact, *TGFBR1* and *TGFBR2* mutations cause Loeys-Dietz syndrome (LDS), which is characterized

by arterial aneurysms, tortuous arteries, Marfanoid habitus, and craniofacial features such as hypertelorism, bifid uvula, and/or cleft palate [Loeys et al., 2006].

Thanks to these initial steps more than 1000 *FBN1* and 100 *TGFBR1/2* mutations have been identified so far (Human Gene Mutation Database, HGMD® Professional 2011.2) and our knowledge about the molecular basis and pathogenesis of Marfan syndrome has remarkably increased during this time.

1.2 Clinical Symptoms

1.2.1 Skeletal System

In patients with classical Marfan syndrome (MFS1) the most visible clinical signs manifest in the skeletal system. Individuals with MFS1 grow typically taller compared to the average and display long slender limbs (dolychostenomelia) and fingers (arachnodactyly). In addition, the abnormal indentation (pectus excavatum) or protrusion (pectus carinatum) of the sternum is characteristic of MFS1 as well (Figure 2).

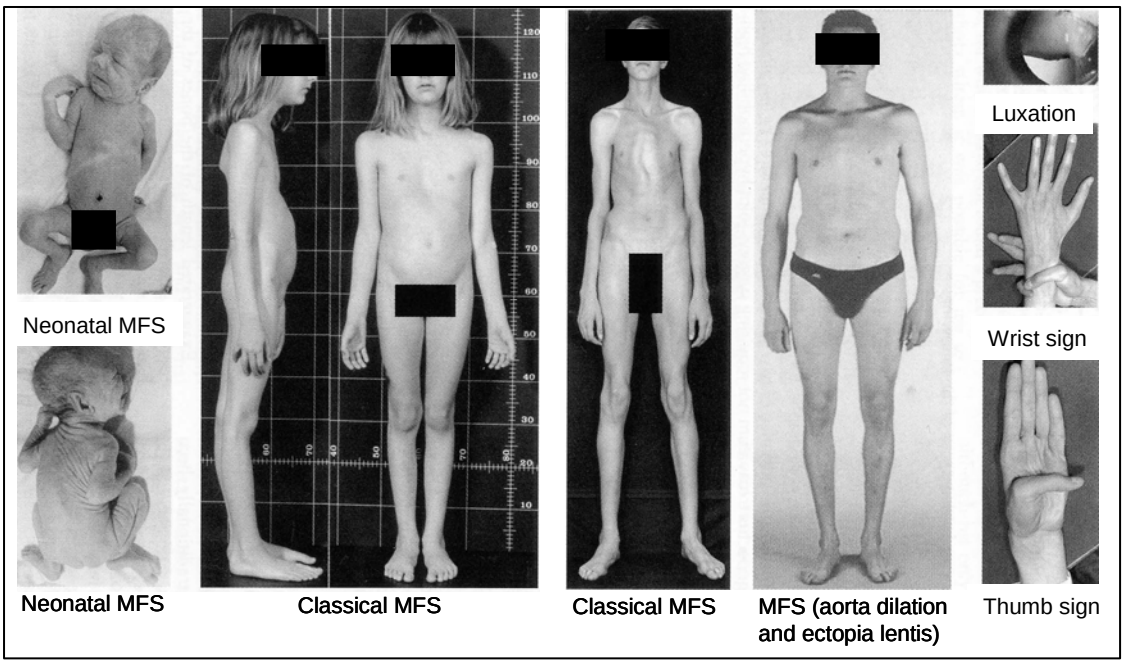

Figure 2. Selected examples showing clinical manifestations of Marfan syndrome (MFS) [Royce and Steinmann (eds). 2002. Connective Tissue and Its Heritable Disorders, p. 588].

The abnormal curvature of the spinal cord (scoliosis) affects about 60% of Marfan patients, which progresses rapidly during growth, leading to back pain and ventilatory difficulties. Dilatation of the dural sac (dural ectasia) is present in about 63-92% of patients and seems to be the causing effect of back pain, headache, weakness, and loss of sensation above and below the affected limb, occasional rectal pain, and pain in the genital area [Altman et al.,

2008]. Other signs, as joint hypermobility, high-arched palate, and flat feet, are also commonly associated with the syndrome [Dean, 2007].

1.2.2 Cardiovascular System

The main cause of premature death of patients with Marfan syndrome is the progressive enlargement of the aortic root leading to sudden dissection. The aortic wall is a flexible formation of elastic fibers produced by smooth muscle cells and fibroblasts. Due to abnormal connective tissue, the wall of the aorta is weakened and can stretch out, causing the dilation of the blood vessel (aneurysm) [Milewicz et al., 2005]. The average diameter of the ascending aorta in adult males is 3.6-3.8 cm. Patients have a risk for aortic dissection if the diameter of the aneurysm of the aortic sinus (sinus of Valsalva) is greater than 5 cm, or the dilatation increases more than 5% per year, or 1.5 mm/year in adults. But other cardiovascular complications, such as mitral valve prolapse (MVP) and mitral regurgitation (MR), pulmonary artery dilatation, and aortic valve incompetence due to dilatation of the aortic root, can be typical for the syndrome too [Treasure, 2000].

1.2.3 Ocular System

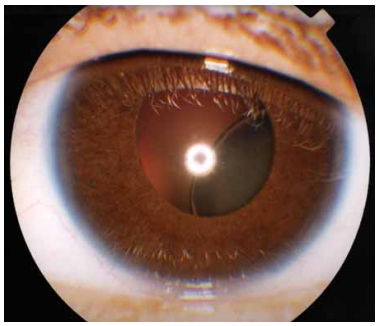

Figure 3. Lens subluxation (http://www.medstudens.com.br/original/revisao/marfan/marfan.htm).

Characteristic ocular features of Marfan syndrome include ectopia lentis (50-80%), myopia (34-44%), and retinal detachment (5-11%). Lens dislocation (ectopia lentis) is characterized by the displacement or malposition of the lens (Figure 3). In the case of Marfan syndrome, ectopia lentis is usually bilaterial and develops in early childhood or in the second decade. The second prominent ocular manifestation is myopia, which is also called nearsightedness or shortsightedness, and is the refractive defect of the eye. Myopia is associated with increased ocular axial length and increased risk of retinal detachment [Judge and Dietz, 2005; Dean, 2007; Nahum and Spierer, 2008].

1.2.4 Respiratory System

Severe pectus excavatum or progressive scoliosis can evoke abnormalities of the pulmonary function in patients with Marfan syndrome. Spontaneuos pneumothorax was found in 4-11% of Marfan patients. The abnormal connective tissue can cause the formation of tiny air sacs making the lungs less elastic. However, Marfan patients generally do not experience noticeable problems with their lungs, but if these tiny air sacs become stretched or

swollen, the risk of lung collapse may increase [Streeten et al., 1987; Judge and Dietz, 2005; Dean, 2007].

1.2.5 Central Nervous System

The dilatation of the dural sac (dural ectasia) is sometimes associated with intracranial hypotension-associated headache. But weakness, back pain, rectal pain, and pain in the genital area have also been described [Altman et al., 2008].

1.3 Molecular Genetics of Marfan Syndrome

1.3.1 The *FBN1* Gene

In 1990, Marfan syndrome was mapped to chromosome 15 in five families using linkage analyses with polymorphic markers [Kainulainen et al., 1990]. One year later, the locus was further mapped and the first mutation in the gene fibrillin-1 (*FBN1*) was identified [Dietz et al., 1991a; Dietz et al., 1991b]. The gene is located on the long arm of chromosome 15 at position q21.1 and contains 65 exons with large intronic regions. The coding sequence has a length of 8613 bp. Upstream of exon 1 (previously termed as exon M), which contains the translation initiating codon, three alternatively spliced exons were found (B, A, and C). These alternatively spliced exons generate some ambiguity regarding the translation initiation [Corson et al., 1993]. High evolutionary conservation of the 5`UTR sequences suggested a regulatory role for this region [Biery et al., 1999]. Through DNA sequence alignment of the human and six additional mammalian species of this region several highly conserved sequence blocks were discovered: 1. a 66 bp ultraconserved sequence, containing an initiator element, 2. a downstream promoter element, and 3. a 10 bp palindromic element. In addition, reporter constructs containing part of the 5` upstream exons and introns showed increased luciferase activity as well. These results suggest multiple promoter sequences in these regions [Guo et al., 2008]. Recently, the alternatively spliced region was analyzed bioinformatically and compared to those in non-human primates in order to find conserved vertebrate transcription factor binding sites. High homology and conservation was found, suggesting a functional role of this region [Singh et al., 2008]. Summers and colleagues characterized the 5` region of *FBN1* for both mouse and human experimentally and bioinformatically and found a highly conserved single CpG-rich promoter in exon A [Summers et al., 2009]. The expression pattern of fibrillin-1 is ubiquitous. However, there is a quantitative difference in the amounts of *FBN1* transcripts in different tissues [Magyar et al., 2009].

The *FBN1* mRNA is translated into the approximately 350-kDa extracellular glycoprotein profibrillin-1, which is secreted and proteolytically processed at the N– and C– termini [Sakai et al., 1986; Milewicz et al., 1992; Raghunath et al., 1994; Raghunath et al., 1995; Reinhardt et al., 1996; Lönnqvist et al., 1998; Ritty et al., 1999]. The mature fibrillin-1 protein is about 320 kDa and comprises the major component of the 10–12 nm calcium-binding microfibrils [Sakai et al., 1991]. The fibrillin-1 protein contains two repeated domains, the epidermal growth factor-like (EGF) domain and the transforming growth factor-1 binding protein-like (TB) domain [Corson et al., 1993; Pereira et al., 1993]. 43 of the 47 EGF domains are associated with calcium binding, while the seven TB domains are distributed throughout fibrillin-1 and separated by a variable number of tandemly repeated EGF domains [Giltay et al., 1999]. Each TB domain contains eight cysteins, which are predicted to form four intramolecular disulphide bonds [Yuan et al., 1997].

The distribution of different classes of *FBN1* mutations (collected in the locus-specific mutation database (UMD, http://www.umd.be) is shown in Figure 4.

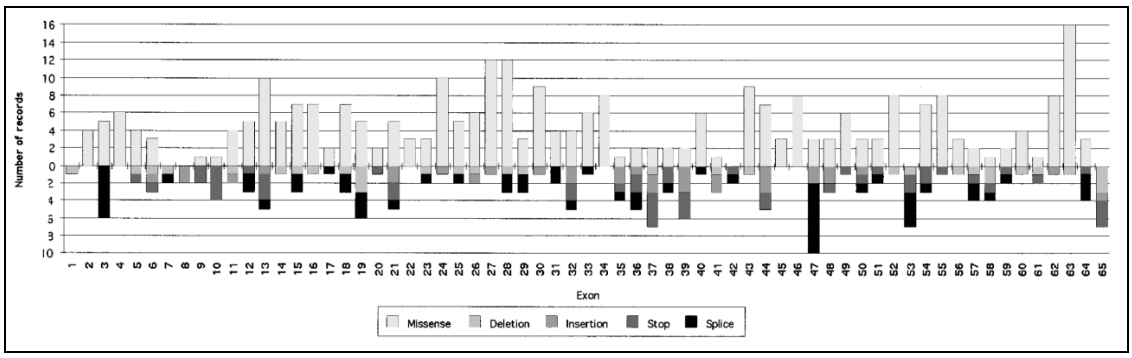

Figure 4. Distribution of *FBN1* mutations reported in UMD. The x-axis represents the 65 exons of the *FBN1* gene and the y-axis displays the number of entries [Collod-Beroud and Boileau, 2002].

In our cohort of ~400 unrelated patients, we have detected 135 *FBN1* sequence variants, comprising missense (55%), nonsense (12%), and splice site mutations (18%) as well as small deletions (10%), insertions (1%), and insertion-deletions (4%). Some genotype-phenotype correlations among mutation types and clinical manifestations have been described [e.g. Faivre et al., 2007], especially for neonatal MFS, which is associated with mutations in exons 24–32 [Collod-Beroud and Boileau, 2002]. To date, 17 cases of large DNA rearrangements have been found [e.g. Kainulainen et al., 1992; Liu et al., 2001; Loeys et al., 2004; Singh et al., 2006a; Mátyás et al., 2007; Faivre et al., 2010; Hilhorst-Hofstee et al., 2011].

1.3.2 *TGFBR1* and *TGFBR2* Genes

Heterozygous mutations in the genes encoding *TGFBR1* and *TGFBR2* have been identified in MFS2, LDS, and familial thoracic aortic aneurysms and dissections (TAAD), indicating genetic heterogeneity in MFS and its related disorders [e.g. Mizuguchi et al., 2004; Loeys et al., 2005; Pannu et al., 2005; Loeys et al., 2006; Mátyás et al., 2006].

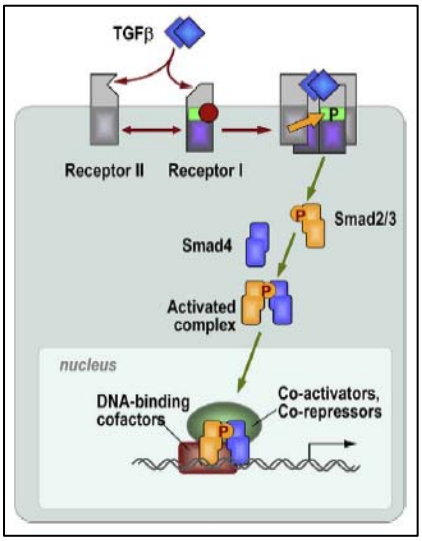

Figure 5. The TGFβ canonical signaling pathway [Massague and Gomis, 2006].

The *TGFBR1* gene is located on chromosome 9q22, spans a genomic size of about 49 kb and contains 9 exons with an open reading frame (ORF) of 1512 bp. *TGFBR2* was mapped to chromosome 3p24.1, containing a genomic length of approximately 88 kb and 8 exons with an ORF of 1779 bp.

TGFBR1 and TGFBR2 are Ser/Thr kinase receptors with similar structure (Figure 5). Both proteins have a small Cys-rich extracellular domain, a single transmembrane spanning-region and an intracellular part with a kinase domain [ten Dijke and Arthur, 2007]. During canonical signaling, TGFβ binds to the dimerized TGFBR2, recruiting the dimerized TGFBR1 and thus forming a receptor-ligand complex. Immediately, TGFBR2 phosphorylates the serine/threonine residues in the regulatory region (GS region) of TGFBR1, thereby activating it. TGFBR1 specifically phosphorylates receptor-bound SMAD molecules (SMAD2/3) at its carboxy-terminal SXS motif, allowing the release from the receptor and the binding of SMAD4 to this region. This SMAD complex translocates into the nucleus, where, together with different nuclear factors, it modulates the trancription of target genes [Massague and Gomis, 2006; Ramirez and Rifkin, 2009].

TGFBR1 and TGFBR2 are involved in several cellular processes, including growth inhibition, apoptosis, proliferation, and extracellular matrix production, but their function in the pathogenesis of MFS/LDS is poorly understood.

1.4 Pathogenesis of Marfan Syndrome

1.4.1 MFS and Extracellular Microfibrils

The 10–12 nm microfibrils are found in elastic as well as in non-elastic tissues with multiple structural functions, such as maintenance of elastic fibers, anchoring epithelial cells to the interstitial matrix, and formation and maintaining homeostasis of the elastic matrix

[Mecham and Heuser, 1990; Pereira et al., 1997]. The loss of aortic integrity was explained as a consequence of altered structural integrity of microfibrils [Byers, 2004]. This model is called dominant-negative effect, in which mutant and wild-type fibrillin-1 proteins interfere with each other [Aoyama et al., 1993; Dietz et al., 1993]. In contrast, reports on patients with low levels of mutant fibrillin-1 and classic severe MFS refuse this hypothesis [Halliday et al., 1999]. In this case, the mutant transcript containing a premature termination codon (PTC) is detected by the evolutionary conserved quality control mechanism, called nonsense-mediated mRNA decay (NMD), which prevents the expression of a truncated protein. The process consists of a series of steps that ultimately lead to the degradation of mRNA, resulting in functional haploinsufficiency [Judge et al., 2004; Wilkinson, 2005; Chan et al., 2007; Stalder and Mühlemann, 2008]. Since the identification of the *FBN1* gene and the first mutation, the molecular pathogenesis of the syndrome (i.e. dominant negative versus haploinsufficiency) is the subject of intense discussions. Animal models suggested that functional haploinsufficiency is sufficient to cause features of MFS [Judge et al., 2004]. This hypothesis was confirmed in MFS patients by the identification of large *FBN1* deletions which abolish transcription from the mutant allele, leading to true haploinsufficiency [Mátyás et al., 2007]. Taken together, publications as well as our own results suggest that both pathogenic mechanisms can lead to MFS [Aoyama et al., 1994; Eldadah et al., 1995; Judge et al., 2004; Mátyás et al., 2007; Magyar et al., 2009].

Some features of MFS, such as long bone and rib overgrowth or alterations in muscle growth, cannot be explained by structural defects, suggesting that the role of fibrillin-1 cannot be simplified as a simple structural protein of the extracellular matrix (ECM). Recently, the role of TGFβ signaling was demonstrated as a major contributor to the pathophysiology of MFS [Neptun et al., 2003].

TGFβ-1, -2, and -3 (hereafter collectively called as TGFβs unless otherwise indicated) are synthesized as a precursor and are proteolytically processed to a mature peptide and a latency associated peptide (LAP). After cleavage, the dimers of mature TGFβs and one LAP remain associated via non-covalent bonds and form the small latent complex (SLC). This complex is biologically inactive. The SLC can covalently attach to TGFβ binding proteins (LTBPs) to form the large latent complex (LLC) [Kaartinen and Warburton, 2003; Byers, 2004; ten Dijke and Arthur, 2007].

The LTBP-1 is associated with fibrillin-1 in the ECM (Figure 6). This protein was found to bind the latent form of TGFβ-1, a multifunctional growth factor with crucial roles in the development and tissue homeostasis [Massague and Gomis, 2006]. In addition, it was

reported that fibrillin-1 itself (encoded by exons 44-49) can release endogenous TGFβ-1 by binding to the N-terminal large latent complex (LLC) interacting region, and thereby stimulating the TGFβ signaling pathway [Chaudhry et al., 2007].

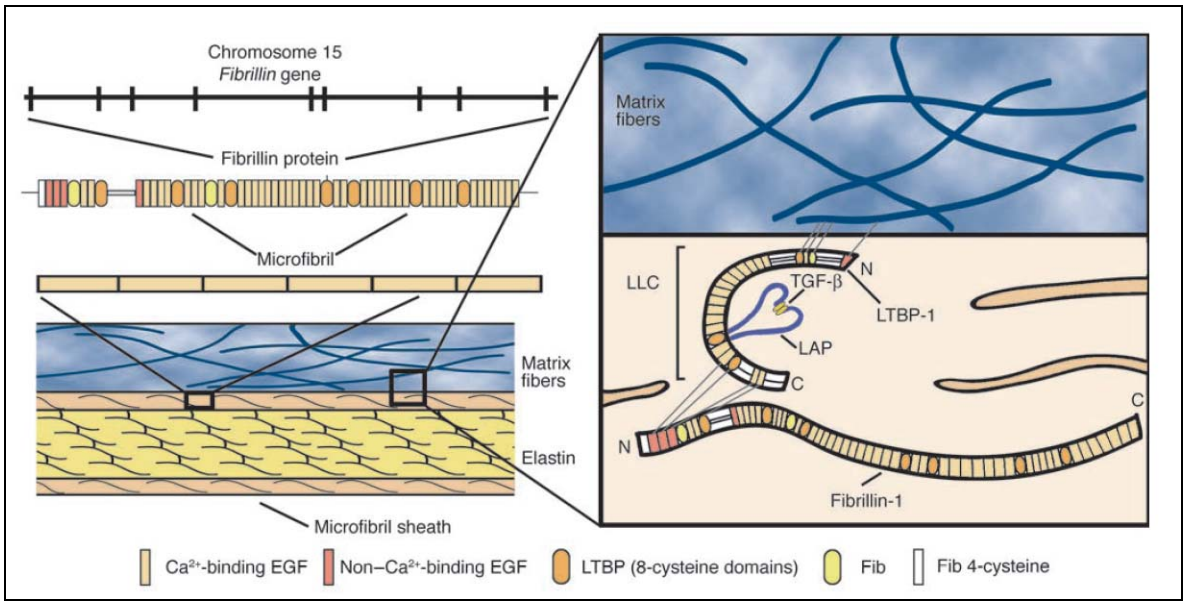

Figure 6. Fibrillin-1 and its interaction with LTBP-1 to control TGFβ [Byers, 2004].

Consequently, fibrillin-1 has not only a structural function but also mediates TGFβ signaling by storing, presenting, and releasing the ligands.

1.5 Diagnosis of Marfan Syndrome

1.5.1 Ghent Nosology

In the last nearly 15 years, the clinical diagnosis of Marfan syndrome was based on the Ghent nosology, which is the revised and most stringent follow-up of the Berlin nosology, defined in 1986 [Beighton et al., 1988; De Paepe et al., 1996; Summers et al., 2006]. This systematic classification discriminates the clinical manifestations of the skeletal, ocular, cardiovascular, and pulmonary systems as well as the dura and skin findings into major and minor criteria. Accordingly, the clinical diagnosis of Marfan syndrome requires the presence of at least two major criteria in different organ systems and the involvement of a third organ system, in the absence of family history. In the presence of an MFS-causing *FBN1* mutation, one major criterion and the involvement of a second organ system is enough for the clinical diagnosis of MFS. In spite of its good specificity, the Ghent nosology has recently been revised [Loeys et al., 2010].

The Ghent nosology was criticized to neglect or insufficiently taking into account the age-dependent manifestations (the diagnosis of children is therefore more difficult) and weighting

poorly validated or non-specific features. The revised Ghent nosology (Table 1) emphasizes two cardinal features, the aortic root aneurysm and ectopia lentis. These manifestations are sufficient for the clinical diagnosis, even if no family history is present. In the absence of any of these cardinal features, a mutation in the *FBN1* gene or the combination of systemic features is required [Loeys et al., 2010; Faivre et al., 2011].

Table 1. Revised diagnostic criteria for Marfan syndrome (Revised Ghent nosology) [Loeys et al., 2010].

In the absence of a family history:
(1) Ao ($Z \geq 2$) AND EL = MFS
(2) Ao ($Z \geq 2$) AND *FBN1* = MFS
(3) Ao ($Z \geq 2$) AND Syst ($\geq$7 points) = MFS[a]
(4) EL AND *FBN1* with known Ao = MFS

EL with or without Syst AND with an *FBN1* not known with Ao or no *FBN1* = ELS Ao ($Z < 2$) AND Syst ($\geq$5) with at least one skeletal feature without EL = MASS MVP AND Ao ($Z < 2$) AND Syst (>5) without EL = MVPS

In the presence of a family history:
(5) EL AND FH of MFS (as defined above) = MFS
(6) Syst ($\geq$7 points) AND FH of MFS (as defined above) = MFS[a]
(7) Ao ($Z \geq 2$ above 20 years old, $\geq$3 below 20 years) + FH of MFS (as defined above) = MFS[a]

Systemic score

- Wrist AND thumb sign – **3** (Wrist OR thumb sign – **1**)
- Pectus carinatum deformity – **2** (pectus excavatum or chest asymmetry – **1**)
- Hindfoot deformity – **2** (plain pes planus – **1**)
- Pneumothorax – **2**
- Dural ectasia – **2**
- Protrusio acetabuli – **2**
- Reduced US/LS AND increased arm/height AND no severe scoliosis – **1**
- Scoliosis or thoracolumbar kyphosis – **1**
- Reduced elbow extension – **1**
- Facial features (3/5) – **1** (dolichocephaly, enophtalmos, downslanting palpebral fissures, malar hyoplasia, retrognathia)
- Skin striae – **1**
- Myopia >3 diopters – **1**
- Mitral valve prolapse (all types) – **1**

Maximum total: 20 points; score $\geq$7 indicates systemic involvement

Ao, aortic diameter at the sinuses of valsalva above indicated *Z*-score or aortic root dissection; EL, ectopia lentis; ELS, ectopia lentis syndrome; *FBN1*, fibrillin-1 mutation; *FBN1* with known Ao, *FBN1* mutation that has been identified in an individual with aortic aneurysm; FH, family history; MASS, myopia, mitral valve prolapse, aortic root dilation, skeletal findings, striae syndrome; MVPS, mitral valve prolapse syndrome; Syst, systemic score; US/LS, upper segment/lower segment ratio; *Z*, *Z*-score.
[a]Caveat: without discriminating features of Shprintzen-Goldberg syndrome, Loeys-Dietz syndrome or vascular EDS syndrome.
– if present, then TGFBR1/2 testing, collagen biochemistry, COL3A1 testing
– other conditions/genes will emerge with time.

1.5.2 Differential Diagnoses

1.5.2.1 Disorders with Overlap to the Skeletal Symptoms of MFS

1.5.2.1.1 Congenital Contractural Arachnodactyly (CCA) [OMIM 121050]

CCA is also called Beals syndrome or Beals-Hecht syndrome after the first description [Hecht and Beals, 1972]. CCA is characterized by a marfanoid habitus, crumpled ears (folded upper helix of the external ear), flexion contractures, severe kyphoscoliosis, and in the majority of affected individuals, muscular hypoplasia (Figure 7). Most of the phenotypic features are associated with Marfan syndrome as well. However, patients with CCA just occasionally have cardiovascular and ocular difficulties [Hecht and Beals, 1972; Robinson et al., 2006; Nishimura et al., 2007]. The syndrome is inherited in an autosomal dominant

manner and is caused by mutations in the fibrillin-2 (*FBN2)* gene. *FBN2* contains 65 exons, is located on chromosome 5q23-31, and encodes a large extracellular matrix protein, designated fibrillin-2 [Putnam et al., 1995].

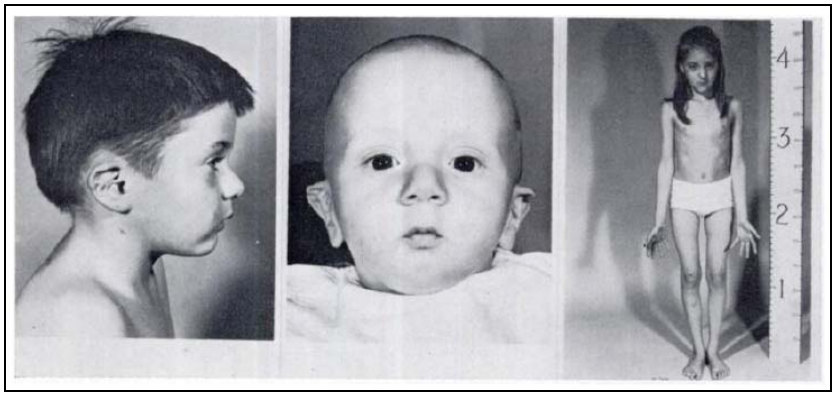

Figure 7. Patients with Congenital Contractural Arachnodactyly (CCA) [Hecht and Beals, 1972].

1.5.2.1.2 Shprintzen-Goldberg Craniosynostosis Syndrome (SGS) [OMIM 182212]

Figure 8. Craniofacial and skeletal features of SGS [Robinson et al., 2005].

The Shprintzen-Goldberg syndrome is characterized by craniosynostosis and other facial features, skeletal changes, neurological changes, cardiovascular and connective tissue abnormalities (Figure 8). There are two characteristic features which are different from Marfan syndrome. First, the involvement of the nervous system with mild to moderate mental retardation, hydrocephalus, and dilatation of the lateral ventricles. Second, the cardiovascular difficulties, by which aortic root dilatation is most likely not found. The molecular basis of the syndrome has not been elucidated. It has been shown that mutations in the *FBN1* or *TGFBR2* genes are associated with SGS [Robinson et al., 2005; Kosaki et al., 2006].

1.5.2.2 Disorders with Overlap to the Cardiovascular Symptoms of MFS

1.5.2.2.1 Familial Thoracic Aortic Aneurysms and Dissections (TAAD) [OMIM 607086, 608967]

Familial aggregation studies have indicated that up to 20% of nonsyndromic aortic aneurysm (Figure 9) and/or dissection (AAD) patients have a familial background [Morisaki et al., 2009]. This type of AAD was shown to be inherited primarily in an autosomal dominant manner with decreased penetrance and variable expression of the disease symptoms [Pannu et al., 2005].

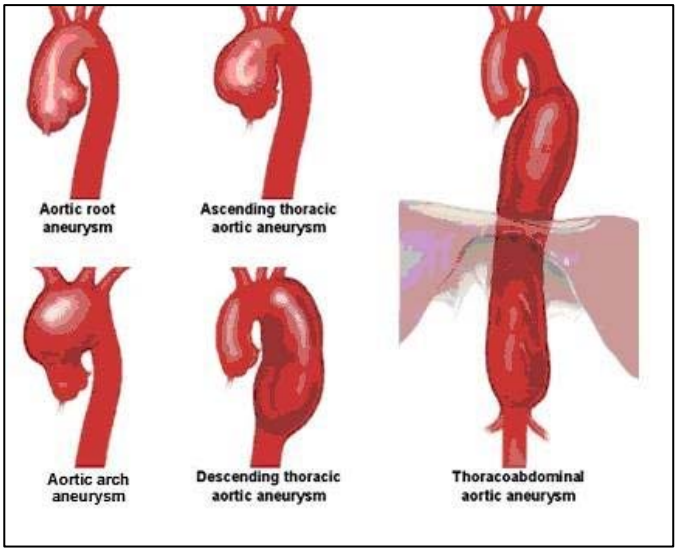

Figure 9. Different types of aortic aneurysms (https://secure.mawebcenters.com/websites/cardiacandthoracicsurgicalassociates/ThoracicAorticAneurysm.html).

Several genetic loci are linked to TAAD: the *TAAD1* locus was mapped to chromosome 5q13-q14, the *FAA1* to 11q23-q24, and the *TAAD2* to 3p24-p25 [Guo et al., 2001; Vaughan et al., 2001; Hasham et al., 2003]. Interestingly, the gene encoding the TGFBR2 protein was found to be mutated in patients with TAAD2, which was previously described in association with MFS2 and was located on chromosome 3p24-p25. Recently, mutations in the genes *TGFBR1*, *MYH11*, *ACTA2* and *MYLK* have also been identified in patients with TAAD [Mátyás et al., 2006; Zhu et al., 2006; Guo et al., 2009; Wang et al., 2010].

1.5.2.2.2 Mitral Valve Prolapse Syndrome [OMIM 157700]

Mitral valve prolapse (MVP) is a common cardiac disorder caused by the uneven closure of the valve`s leaflets during each heartbeat. This is because of the valves are too large or the heart strings are too long. In spite of these abnormalities, about 60% of people with MVP have no symptoms over the time, but they are more susceptible to get infective endocarditis and increased risk of stroke and sudden death (www.emedicinehealth.com/mitral_valve_prolapse/article_em.htm). MVP may occur with greater frequency in patients with connective tissue disorder like MFS or Ehlers-Danlos syndrome. Although the association of MVP with MFS is likely, the link to *FBN1* was excluded. Genome wide linkage analysis mapped MVP to chromosomes 16p11.2-p12.1, 11p15.4, and 13q31.3-31.2 [Levine and Slaugenhaupt, 2007].

1.5.2.2.3 MASS Phenotype [OMIM 604308]

The MASS phenotype covers remarkable similarities with MFS. The name MASS derives from the first characters of following features: **M**itral valve prolapse, **A**ortic root dilatation, **S**kin stretch marks, and **S**keletal features. The only difference is that the eyes are not affected. The disorder is inherited in an autosomal dominant manner and caused by mutations in the *FBN1* gene [Dietz et al., 1993].

1.5.2.2.4 Loeys-Dietz Syndrome (LDS) [OMIM 609192, 610380, 608967, 610168]

LDS is characterized by widely spaced eyes, bifid uvula, cleft palate, and arterial tortuosity with ascending aortic aneurysm and dissection. The syndrome is inherited in an autosomal dominant manner and displays variable clinical expression [Loeys et al., 2005]. There are two genes associated with LDS, the *TGFBR1* and *TGFBR2*. Interestingly, mutations in *TGFBR1* and *TGFBR2* were also found in patients showing arterial tortuosity with ascending aortic aneurysm and dissection, but without craniofacial abnormalities. Based on the presence or the absence of craniofacial involvement, the syndrome was subdivided in LDS type 1 (LDSI) and 2 (LDSII) [Loeys et al., 2006].

1.5.2.3 Disorders with Overlap to the Ocular Symptoms of MFS

1.5.2.3.1 Homocystinuria [OMIM 236200]

Homocystinuria, also known as cystathionine β-synthase deficiency, is an autosomal recessive, multisystemic disorder with manifestations in the ocular, skeletal, cardiovascular, and nervous system. Disease-associated mutations have been identified in the *CBS* gene, which encodes the cystathionine β-synthase enzyme [Moat et al., 2004; Skovby et al., 2010].

1.5.2.3.2 Weill-Marchesani Syndrome [OMIM 608328]

The Weill-Marchesani syndrome is a rare connective tissue disorder characterized by short stature, short and stubby hands, and occasionally joint stiffness. In most of the cases the ocular system is also affected, which can include microspherophakia, myopia, ectopia lentis, glaucoma, and cataract. Both autosomal dominant and autosomal recessive inheritance have been reported. In association with the autosomal recessive form, mutations in the *ADAMTS10* gene have been described, which is localized on chromosome 19p13.3 and encodes a metalloproteinase with a thrombospondin motif [Faivre et al., 2003a]. In one family, a mutation in the *FBN1* gene has been identified with autosomal dominant inheritance [Faivre et al., 2003b].

1.5.2.3.3 Ectopia Lentis (EL) [OMIM 129600]

Isolated EL is a genetically heterogenous condition defined by characteristic displacement of the lens and caused by the disruption of the zonular fibers (Figure 3). In the majority of the cases EL is attributed to systemic diseases, such as MFS, homocystinuria, and Weill-Marchesani syndrome. However, EL can occur as an isolated form with autosomal dominant or autosomal recessive inheritance. Mutations in the *FBN1* gene have been described to be associated with the autosomal dominant form, while mutations in *ADAMTSL4* were recently reported to cause autosomal recessive EL [Vanita et al., 2007; Ahram et al., 2009].

1.5.2.3.4 Stickler Syndrome [OMIM 108300, 604841, 184840]

Stickler syndrome is an inherited connective tissue disorder with characteristic ophthalmic, orofacial, auditory, and articular manifestations. The syndrome is subclassified in four types, which are distinguished by their genetic origin and characteristic symptoms. Four collagen genes have been described in association with the syndrome: *COL2A1*, *COL9A1*, *COL11A1*, and *COL11A2* [Snead and Yates, 1999]. These genes are involved in the production of three collagen types, type II, type IX, and type XI. *COL2A1*, *COL11A1,* and *COL11A2* are involved in the autosomal dominantly inherited form, while *COL9A1* was described in association with autosomal recessive inheritance [Snead and Yates, 1999; Van Camp et al., 2006]. Collagens are major macromolecules that provide structure and strength to connective tissues and therefore support the body's joints and organs.

1.6 Pre-mRNA Splicing

The majority of mammalian genes contain relatively short exons (50-250 base pairs) that are separated by much longer introns (hundreds to thousands of base pairs). To produce correct functional proteins introns are removed from the RNA transcripts and exons are accurately joined together by the splicing process. The accurate recognition of the exon-intron boundaries requires *cis*-acting elements, such as 5` splice sites (donor sites), 3` splice sites (acceptor sites), branch sites, and polypyrimidine tracts (Figure 10).

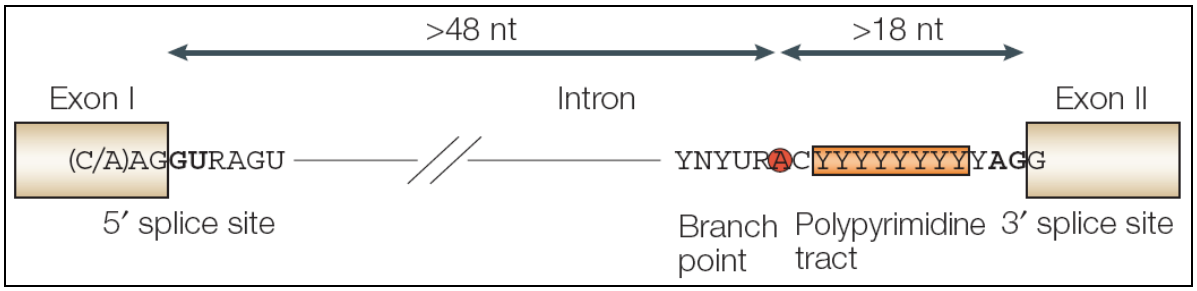

Figure 10. Important pre-mRNA sequences for splicing [Matlin et al., 2005].

The human splice donor site includes the last 3 nucleotides of the upstream exon and the first 6 nucleotides of the intron (positions -3 to +6). These sequences are recognized by the U1 small nuclear ribonucleoprotein (snRNP) via base-pairing. For human GT-AG introns the 5`splice site consensus sequence is: MAG|GURAGU (M = C or A, R = purine; | denotes the exon-intron border). The splice acceptor site is more complex than the donor site. It can be subdivided into three parts: (1) a less conserved branch point sequence YNYURAY (Y = pyrimidine, R = purine, N = any nucleotide, the branch point is underlined), (2) the polypyrimidine tract, (3) and the invariant 3` splice site. The branch point is located ~18-40 nucleotides from the 3` AG, but can also be placed deeper in the intron. The polypyrimidine stretch composes a long uridine tail which can be interrupted by cytosines or purines. The 3` splice site is the first recognized AG dinucleotide downstream of the polypyrimidine tract [Hartmann et al., 2008].

These donor and acceptor splice site signals alone, however, are not enough for accurate splicing. Additionally, short exonic and intronic elements are needed for correct splice site recognition, which can act by stimulating (enhancer motifs) or repressing splicing (silencer motifs).

1.6.1 Splicing Enhancer and Silencer Motifs

Exonic splicing enhancers (ESEs) are high purine containing 6-8 bp sequences that are recognized by serine/arginine-rich (SR) proteins. These splicing factors are characterized by

one or two N-terminal RNA recognition motifs (RRMs) and a C-terminal serine/arginine – rich (RS) domain. The RRM domain is able to bind to RNA by its sequence-specific part, whereas the RS domain has an important role in protein-protein interaction by recruiting spliceosomal components to splice sites and thereby promote exon definition. However, other models described an RS domain independent way, which suggests that SR proteins enhance splicing by antagonizing adjacent silencer elements. The two models are not mutually exclusive [Cartegni et al., 2002; Solis et al., 2008].

SR proteins act primarily in the early stage of spliceosome formation. They promote the binding of U1 snRNP to the 5` splice site and the U2 protein to the pre-mRNA branch site. But they also promote the B complex formation by facilitating the recruitment of U4/U6 and U5 snRNPs [Blencowe, 2000]. More than 20 members of the SR protein family have been described so far [Fu, 1995].

In addition, exons also contain exonic splicing silencer (ESS) motifs which can inhibit pre-mRNA splicing. It is important to note, that the same sequence motif can serve as ESE or ESS, depending on its position with respect to the splice site. Moreover, intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs) can also influence the process of splicing [Hartmann et al., 2008].

1.6.2 Spliceosome Assembly

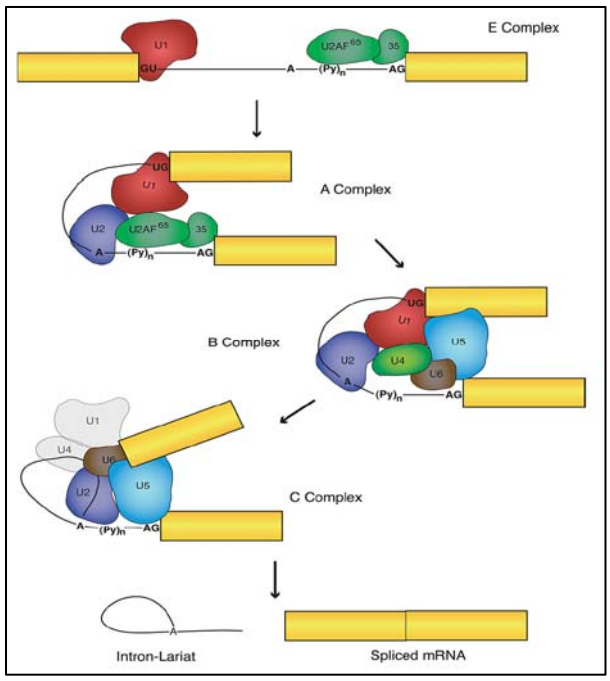

Figure 11. Spliceosome assembly [Solis et al., 2008].

The accurate splicing mechanism is regulated by the splicing machinery, which is a complex of snRNPs. These RNA-protein complexes bind to *cis*-acting elements that help to recognize splice site sequences, and assembly into complexes. In the first step (E complex), the consensus sequences and the exonic/intronic enhancer motifs are recognized. The U1 protein binds to the 5` donor site, the 65 kDa subunit of the U2AF protein (U2AF65) interacts with the polypyrimidine tract, while the 35 kDa subunit (U2AF35) binds to the 3` splice site. The SF1 (splicing factor 1) protein binds to the branch point sequence and

ESEs are recognized by SR proteins. In the next step (A complex), the U2 protein is recruited by the U2AF65 and binds to the branch point. The B complex involves the binding of U4/U6-U5 to the machinery, which, through interaction with other snRNPs and the target RNA, cause the rearrangement of the complex. Subsequently (C complex), the U4/U6 interaction is disrupted, which allows U6 to bind to the 5` splice site and to interact with U2. This disruption causes the displacement of U1 and U4. The U5 protein is a bridge between the donor and acceptor site, helping to align exons for the ligation reaction (Figure 11).

1.6.3 Splicing Reaction

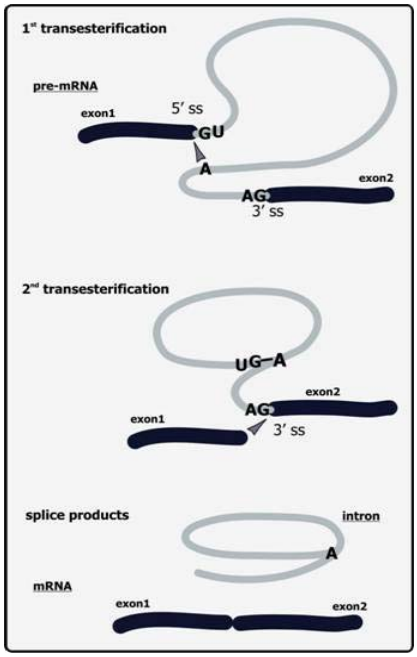

Figure 12. The splicing reaction (http://www.eurasnet.info/alternative-splicing/what-is-alternative-splicing/biochemistry-of-splicing)

The splicing reaction itself contains two trans-esterification reactions (Figure 12). In the first reaction, the intronic 5` splice site phosphate is attacked by the 2` hydroxyl of the branch point adenosine. This attack leads to the cleavage of the upstream exon leaving a free intronic 5` hydroxyl residue, which is ligated to the adenosine of the branch point thereby forming the lariat structure. In the second trans-esterification step, the free 3` hydroxyl of the exon attacks the last nucleotide of the 3` splice site phosphate, finally joining the two exons. In the 5` exon, near the newly formed junction, protein complexes are deposited (called as Exon Junction Complex or shortly EJC). These complexes can have major influence on translation, surveillance, and localization of the spliced mRNA [Cartegni et al., 2002; Matlin et al., 2005; Solis et al., 2008; Wang and Burge, 2008].

1.7 Nonsense-mediated mRNA Decay

Nonsense-mediated mRNA decay (NMD) is an mRNA surveillance mechanism to control the quality of mRNA by eliminating transcripts which contain premature termination codons (PTCs). PTCs can be generated by DNA rearrangements, frameshift or nonsense mutations, errors during transcription, or splicing. All these types of disruptions can lead to deleterious proteins. It has been estimated that NMD regulates ~3-10% of the human and *Drosophila* transcripts [Stalder and Mühlemann, 2008]. Furthermore, it has been reported that ~30% of the known disease associated mutations generate transcripts containing a PTC [Mendell et al., 2004; Stalder and Mühlemann, 2008].

The pre-mRNA splicing generates mature mRNAs that are associated with the cap-binding complex 80 and 20 (CBP80 and CBP20), the poly(A)-binding protein II (PABII), and the EJC (Figure 13). This EJC is located ~20-24 nucleotides upstream of each exon-exon junction and contains a core of proteins including RNPS1, the Y14:MAGOH heterodimer, Barentsz, and RNA helicase eIF4AIII. The EJCs give positional information in the first (pioneer) round of translation by discriminating a premature stop from the natural stop codon. In this first round of translation of normal mRNAs, the stop codon is located downstream from the last EJC and all EJCs can be displaced in this way. After this proofreading step the cap-binding complex is replaced by eIF4E (eukaryotic initiation factor 4E) and PABII is replaced by poly(A)-binding protein I (PABI), and the transcript is proceeded to translation.

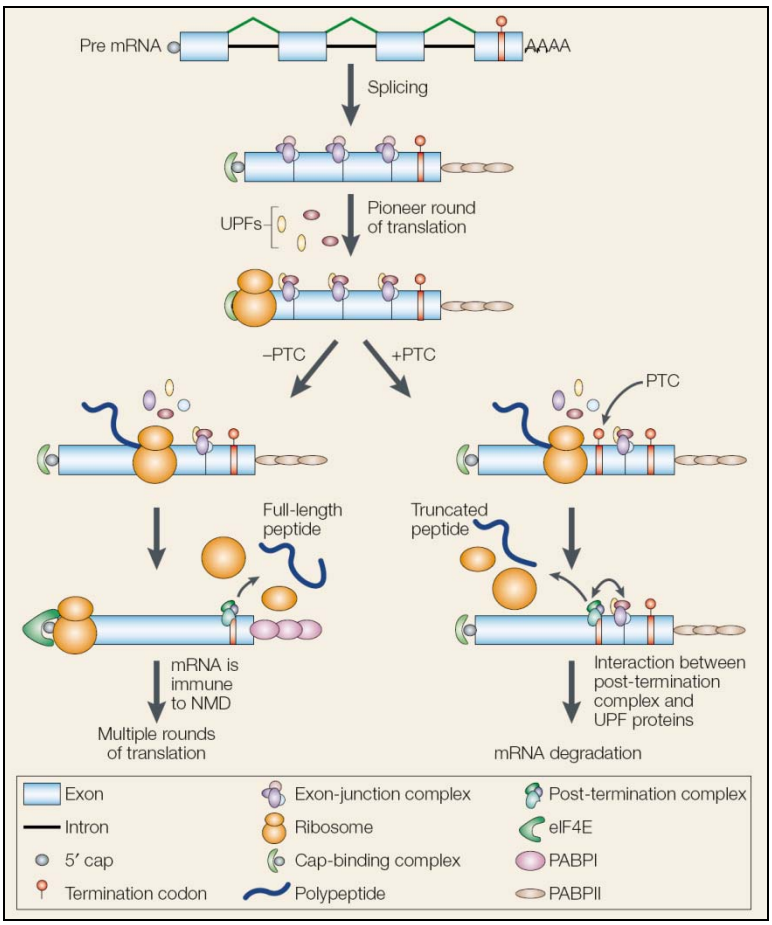

Figure 13. The nonsense-mediated mRNA decay process [Cartegni et al., 2002].

However, when a PTC is recognized more than 50-55 nucleotides upstream of a penultimate exon in the pioneer round of translation, the ribosomes fail to displace the downstream EJC. The exact molecular mechanism for NMD is not yet fully understood. In the popular model of mammalian NMD, the EJC proteins recruit the NMD factor up-frameshift protein 1 (UPF1) if a PTC is recognized. UPF1 is a highly conserved NMD factor,

which interacts with the eukaryotic release factors 1 (eRF1) and 3 (eRF3), binds to the EJC-component up-frameshift protein 2 (UPF2) and subsequently with up-frameshift protein 3b (UPF3b), and further interacts with proteins called suppressor with morphological effect on genitalia 1, 5, 6, and 7 (SMG1, SMG5, SMG6, and SMG7). The key component, the UPF1 protein undergoes a cycle of phosphorylation (by SMG1) and dephosphorylation (by SMG5, SMG6, and SMG7). This phosphorylation cycle is necessary for NMD to occur. The stop codon is recognized by eRF1 and eRF3, which recruit UPF1. This interacts with the protein kinase SMG1 and together forms the so called SURF complex. However, in the presence of a PTC, SURF interacts with the downstream EJC. SMG1 phosphorylates UPF1, leading to subsequent steps of mRNA degradation. In these steps UPF1 is dephosphorylated by SMG5/7, the mRNA looses the cap protein from its 5` end, which makes the RNA susceptible to rapid degradation. In yeast and mammals, the PTC-containing transcripts appear to be degraded by endonucleases from both ends [Holbrook et al., 2004; Chang et al., 2007; Stalder and Mühlemann, 2008].

2 Aim of the Thesis

This thesis aims to add further knowledge to the understanding of the pathogenesis of Marfan syndrome and related diseases. The major aim of my thesis was to investigate the effect of a large number of different exonic and intronic *FBN1* sequence variants on transcript level, with particular interest on pre-mRNA splicing. Furthermore, we used bioinformatic tools for the prediction of splicing aberrations of the above mentioned sequence variants.

In addition, I improved and evaluated Sanger sequencing for quantification of sequence alterations in transcripts and gDNA samples. This approach enabled me to analyze the relative transcript amounts of the *FBN1* gene in fibroblasts and blood samples of patients with suspected Marfan syndrome. Furthermore, I also had the opportunity to apply this method for quantifying relative transcript amounts of *RPGR* and *SLC16A12*.

3 Results

3.1 *FBN1* pre-mRNA Splicing Alterations: Comparative *in vitro* Transcript Analyses and *in silico* Prediction for the Effect of *FBN1* Exonic and Intronic Mutations

Manuscript in preparation

István Magyar,[1] Eliane Arnold,[1] Dvora Colman,[1] Janine Meienberg,[1] Daniela Baumgartner,[3] Armand Bottani,[4] Siv Fokstuen,[4] Marie-Claude Addor,[5] Thierry Carrel,[6] Beat Steinmann,[2] Wolfgang Berger,[1] and Gábor Mátyás[1]

[1] Institute of Medical Molecular Genetics, University of Zurich, Zurich, Switzerland;

[2] Division of Metabolism and Molecular Pediatrics, University Children's Hospital, Zurich, Switzerland;

[3] Department of Pediatric Cardiology, Innsbruck Medical University, Innsbruck, Austria;

[4] Division of Medical Genetics, Geneva University Hospitals, Geneva, Switzerland;

[5] Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland;

[6] Clinic for Cardiovascular Surgery, University Hospital, Berne, Switzerland

Correspondence to:

Gábor Mátyás, Ph.D., Institute of Medical Molecular Genetics, University of Zurich, Schorenstrasse 16, CH-8603 Schwerzenbach, Switzerland; e-mail: matyas@medgen.uzh.ch

Abstract

Pathogenic splicing alterations are increasingly recognized not only as a consequence of mutations in the splice sites but even in exonic *cis*-regulatory elements. The investigation for the pathogenic effect of exonic and deeper intronic variants outside of the canonical splice sites is difficult in the absence of transcript analysis. Thus, web-based prediction tools can be helpful to predict potential splicing defects. Here, we report a total of 93 novel and 32 known sequence variants in the *FBN1* gene from patients with suspected Marfan syndrome, an autosomal dominant connective tissue disorder. These exonic and intronic nucleotide substitutions were characterized by reverse transcription (RT)-PCR and sequence/fragment analyses for the effect on splicing by using RNA either from whole blood and/or fibroblasts. Aberrant transcript processing was found in 27 cases, four of which were induced by exonic variants and 23 by intronic splice site substitutions. Additionally, we tested different exonic splicing enhancer and splice site prediction programs and compared them with our in vitro results. Our data show that *FBN1* exonic mutations can result not only in amino acid changes but also in aberrant splicing. Similarly, splice site mutations can alter splicing in different ways. We found, that splice site web-based tools are very helpful for the prediction of splice site errors; however, the use of ectopic splice sites was not well detected. In contrast, exonic splicing enhancer prediction tools proved to be very weak because just one out of four aberrations was accurately predicted. Hence, transcript analyses are needed to assess the proper consequences of *FBN1* variants at the mRNA level.

Keywords: pre-mRNA splicing; transcript analysis; *in silico* prediction; exonic splicing enhancer; splice site; Marfan syndrome (MFS); fibrillin-1 (*FBN1*)

Introduction

The determination of the human genome and the HapMap projects provided enormous amount of information in the field of human genetics. Today's next-generation sequencing machines are able to determine the entire genetic background in a week and allow simultaneous screening for sequence changes in the patient's genome promising an era of personalized medicine [Mardis, 2008; Tucker et al., 2009]. From these enormous data sets, an important challenge is to identify the variant(s) associated with the disease and to understand the pathogenic effect of those variants.

It appears that approximately 15% of sequence changes can be pathogenic by disrupting pre-messenger RNA (mRNA) splicing [Baralle et al., 2009]. In eukaryotes, pre-mRNA splicing is a complex process by which the non-coding intronic regions are removed and the coding regions are joined together. This sophisticated process needs the interaction of a complex machinery, called spliceosome, and specific signals at both ends of the intron. These signals, the donor site or 5` splice site, the acceptor site or 3` splice site with the branch site, and the polypyrimidine tract sequences; however, are necessary but not sufficient for the accurate operation of this machinery. Additionally, exonic and further intronic *cis*-elements are needed for correct splice site identification, which can enhance or silence splicing by binding *trans*-acting factors. The exonic splicing enhancers (ESEs) are relatively short sequences (6-8 nucleotides) recognized by specific splice factors, the serine/arginine-rich (SR) proteins, exerting positive effect on splice site recognition and stimulating splice site assembly. Exonic and intronic splicing silencers (ESSs/ISSs) and intronic splicing enhancers (ISE) are less well characterized but their role can be as relevant as the exonic enhancer motifs, respectively [Cartegni et al., 2002].

There is increasing evidence that alterations in the coding region can inactivate those exonic *cis*-regulatory motifs and thereby result in aberrant transcript processing. These alterations are originally classified as frame-shift, nonsense, missense or even synonymous. From protein coding viewpoint, frame-shift and nonsense mutation-bearing transcripts undergo nonsense-mediated mRNA decay (NMD) or generate truncated protein isoforms. Missense mutations may influence the structure and function of the protein, whereas synonymous mutations are considered to be neutral.

For routine diagnostic the mutation detection is based on sequencing of genomic DNA (gDNA) as template. However, alterations leading to aberrant pre-mRNA splicing may be impossible to identify by gDNA-based assays [Hartmann et al., 2008]. The latter problem can

be addressed using cDNA analysis. The use of such RNA-based assays is not widespread in routine genetic testing because of practical difficulties in obtaining biopsies as source of RNA. Therefore, *in silico* prediction programs can be helpful in the assessment of these questions [Hartmann et al., 2008; Spurdle et al., 2008; Thusberg and Vihinen, 2009]. A number of prediction programs have been developed to assess the effect of substitutions within the canonical splice sites [Brunak et al., 1991; Reese et al., 1997; Yeo and Burge, 2004; Desmet et al., 2009]. Moreover, recently several computational tools have been designed to predict ESEs in human genomic sequences by searching for hexa- or octanucleotides that are significantly enriched in human exons relative to introns [Cartegni et al., 2002; Cartegni et al., 2003; Fairbrother et al., 2004; Zhang and Chasin, 2004; Zhang et al., 2005].

In this study, we analyzed the effect of numerous exonic and intronic fibrillin-1 (*FBN1*, OMIM 134797) nucleotide substitutions on splicing in patients with suspected Marfan syndrome (MFS, OMIM 154700), an autosomal dominant connective tissue disorder, which displays variable manifestations in the cardiovascular, ocular, and skeletal systems [Dietz et al., 1991; De Paepe et al., 1996]. We report 93 novel and 32 known *FBN1* sequence variants, from which 111 were systematically studied at the RNA-level. In summary, we report on two patients carrying the same silent mutation, two patients harboring missense mutations, one patient with insertion alteration, and 23 patients who carried intronic substitutions. We were able to compare the experimentally determined splicing effects with online *in silico* prediction tools. ESE-prediction appears not very reliable since only one fourth of the observed events were predicted. For intronic substitutions within the canonical splice site the online tools were more reliable, however, the real consequences, whether they cause exon skipping, generate cryptic splice site(s), or even both events, were not well predicted. We discuss the advantages and disadvantages of RNA-analyses and the usefulness of *in silico* prediction tools.

Materials and Methods

Patients and Detection of Mutations

A cohort of 125 patients with MFS-related phenotype was selected for this study including 70 exonic and 55 intronic sequence variants, among which 111 were investigated at mRNA level. Mutation analysis in the coding exons and flanking intronic sequences of the *FBN1* gene was performed as described elsewhere [Mátyás et al., 2002]. Sequence variant nomenclature follows the guidelines of the Human Genome Variation Society (HGVS, www.hgvs.org/mutnomen) with numbering using +1 as A of the ATG start codon of the *FBN1* mRNA reference sequence NM_000138.3. Data on the clinical phenotypes of patients were collected from medical records or during physical examinations by one of the authors (D.B., A.B., S.F., M.-C.A., T.C., and B.S.).

Cell Culture

Fibroblasts from skin biopsies were cultured in Minimal Essential Medium (MEM) with Earle's salts without L-glutamine (PAA Laboratories GmbH, Cölbe, Germany; www.paa.com) supplemented with 10% fetal bovine serum (PAA Laboratories GmbH, Cölbe, Germany; www.paa.com), 1× antibiotic-antimycotic (Gibco, Invitrogen Corporation, Paisley, UK; www.invitrogen.com), and 2 mM L-glutamine (LabForce AG, Nunningen, Switzerland; www.labforce.ch) and incubated at 37°C with 5% CO_2. Prior to RNA extraction, each cell culture at ~100% confluence was split and after they reached the ~80% confluence one subculture (+CHX) was treated with the translation-inhibitor cycloheximide (Actidione; Sigma, St. Louis, MO; www.sigmaaldrich.com) for 6-8 hours at a concentration of 100µg/ml in order to inhibit NMD, while to the other subculture (-CHX) no cycloheximide was added.

DNA and RNA Extraction

Total DNA was extracted from fibroblasts and EDTA-anticoagulated whole blood samples by standard procedures or was referred to us for mutation detection in *FBN1*.

Total RNA was isolated from both cycloheximide-treated and untreated fibroblast subcultures by using the RNeasy mini kit (Qiagen, Hilden, Germany; www.qiagen.com), or the NucleoSpin RNA II kit (Macherey-Nagel AG, Oensingen, Switzerland; www.mn-net.com), and/or from PAXgene-stabilized whole blood using the PAXgene blood RNA kit (Qiagen, Hilden, Germany; www.qiagen.com). Total RNA of Patient 412 was purified from

~30 mg frozen aortic tissue by using parallel the IKA Ultra-Turrax T8 homogenizer and liquid nitrogen to disrupt and homogenize the tissue. Downstream steps of RNA extraction were performed by using the NucleoSpin RNA II kit. DNA and RNA concentrations were determined by using a NanoDrop ND-1000 system (NanoDrop Technologies, Wilmington, DE; www.nanodrop.com) and RNA qualities (RNA integrity number, RIN) were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA; www.agilent.com) according to the manufacturer's instructions.

Transcript Analyses

In order to investigate the effect of mutations on pre-mRNA splicing qualitative transcript analysis was performed. Reverse transcription (RT)-PCR by means of the Qiagen OneStep RT-PCR kit (Qiagen, Hilden, Germany; www.qiagen.com) was used to amplify complementary DNA (cDNA) fragments containing the mutation. RT-PCR, purification of the fragments, and sequencing were performed as described recently [Magyar et al., 2009]. Briefly, RT-PCR products were purified using ExoSAP-IT (USB Corporation, Cleveland, OH; www.usbweb.com). Purified RT-PCR products were sequenced directly in both directions on an ABI PRISM 3100 Genetic Analyzer (ABI3100; Applied Biosystems, Rotkreuz, Switzerland; www.appliedbiosystems.com) and subsequently purified using Sephadex G-50 according to the manufacturer's recommendations (Millipore Technical Note TN053; www.millipore.com). Each exonic mutation was investigated by quantitative sequencing at least once in both directions in order to determine the allele ratios at transcript level. Decreased (>5%, error of the method) or absent mutant allele expression could mean aberrant splicing, while 1:1 allele ratio means intact transcript. Moreover, to detect transcripts of abnormal length, RT-PCR products were separated and visualized by standard gel electrophoresis, as well.

The detected multiple fragments of complex splicing defects were cut out from agarose gel after standard electrophoretic separation for 1-2 hours at 7.7 V/cm. Gel slices were weighted and purified by the Qiagen QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany; www.qiagen.com) eluating the isolated bands in 30 µl Buffer EB. Purified fragments were sequenced directly in both directions as described above.

Fragment Analysis with FAM-labeled Primer

After OneStep RT-PCR with FAM-labeled forward primer (Microsynth AG, Balgach, Switzerland; www.microsynth.ch), 1 µl of the amplified RT-PCR products were mixed with 0.5 µl ROX-500 internal-lane size standard (Applied Biosystems, Rotkreuz, Switzerland;

www.appliedbiosystems.com) and with 10.5 µl Hi-Di formamide (Applied Biosystems, Rotkreuz, Switzerland; www.appliedbiosystems.com) in a total volume of 12 µl were denatured at 95°C for 3 min and for 10min on ice. Fragments were separated by capillary electrophoresis on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Rotkreuz, Switzerland; www.appliedbiosystems.com) using 1sec electrophoretic injection and 15 kV electric field. Data were analyzed using the GenScan Analysis software version 3.1 (Applied Biosystems, Rotkreuz, Switzerland; www.appliedbiosystems.com).

In silico Prediction

All exonic variants were analyzed for ESE disruption by using ESEfinder 3.0 (http://rulai.cshl.edu/tools/ESE), RESCUE-ESE (http://genes.mit.edu/burgelab/rescue-ese), and PESX (http://cubweb.biology.columbia.edu/pesx/). For the prediction of exonic variants the whole exons were analyzed. ESEfinder is a nucleotide-frequency matrix based tool, which allows the prediction of putative ESEs recognized by the human SR proteins SF2/ASF, SC35, SRp40, and SRp55. The output of the program is a series of high scores above the threshold according to the predicted ESE motifs. A high score of the appropriate motif means exon inclusion, while the loss of this score leads to reduced level of exon inclusion [Cartegni et al., 2003]. The second computational approach, the RESCUE-ESE program recognizes hexanucleotides, which are significantly enriched in exons compared to introns and are significantly more frequent in weak splice site-containing exons than in strong ones. The result of the prediction shows the query sequence with the ESE-motifs drawn above in yellow. The disruption of this hexamer suggests altered splicing [Fairbrother et al., 2004]. The PESX web-interface predicts enhancers and silencers based on overrepresentation of octamers that were found more frequently in exons than in pseudo exons or introns. The output displays the input sequence with the ESEs in green and the ESSs in red [Zhang and Chasin, 2004; Zhang et al., 2005].

All intronic alterations were predicted using NNSPLICE V0.9 (http://www.fruitfly.org/seq_tools/splice.html), NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/), Human Splicing Finder 2.4 (http://www.umd.be/SSF/), and ESEfinder 3.0 (http://rulai.cshl.edu/tools/ESE). For all *in silico* analyses the default thresholds were used. For analyses of the intronic variants 600-650 bps from the affected introns and the whole exons were analyzed. Additionaly, all exonic variants were analyzed with the above mentioned splice site programs, in order to predict cryptic splice site generation. The HSF is the updated version of the Splicing Sequences

Finder (SSF), and was developed to predict the effects of variants on splicing and to identify enhancer and silencer motifs. HSF also includes the MaxEntScan script, which is based on the maximum entropy distribution and provides the least biased approximation for the distribution of short sequence motifs. The NNSplice web tool was developed as described by Brunak et al (1991) with the only difference to use genes for training, which have constraint consensus splice sites. The output is a score for potential splice sites. The NetGene2 uses both splice signals and coding information compared to NNSplice, and gives a score for putative splice site. The splice site prediction function of ESEfinder 3.0 is a novel tool. The matrices are based on data of the database of classified alternative splicing events (dbCASE). The threshold values correspond to the first quantile of all splice site scores, and the scores above the threshold display putative splice site.

All exonic and intronic variants were also checked by the ESE (ESEfinder 3.0, RESCUE-ESE) and splice site (NNSplice, MaxEntScan) prediction modules of the Alamut Interpretation Software (http://www.interactive-biosoftware.com/).

Results

Because not only intronic but also exonic variants can induce aberrant splicing by disrupting *cis*-regulatory elements, such as exonic splicing enhancers or silencers (ESEs or ESSs), we performed transcript analyses for a large cohort of *FBN1* nucleotide substitutions. In this study, 52 missense, 12 synonymous, 1 insertion, 4 small exonic deletions, 1 indel, and 55 intronic sequence variants (Table 1) were identified at genomic DNA level (gDNA) and analyzed *in silico* by different ESE status or splice site prediction programs. Further, from this large number of sequence variants 70 exonic and 41 intronic changes were investigated at mRNA level, amplifying the transcript by one-step RT-PCR, following Sanger sequencing.

In vitro and *in silico* Analyses of *FBN1* Exonic Variants

As summarized in Table 2, the c.4192G>A (p.D1398N; Patient 117) and c.1960G>A (p.D654N; Patient 412) missense mutations, the c.6354C>T (p.I2118I; Patients 58 and 245) silent mutation in two unrelated patients, and the c.1981_1982insT insertion (p.C661L*fs*X20; Patient 262) out of 70 exonic variants were detected that associate with aberrant RNA processing. The cDNA analysis of the c.4192G>A missense alteration (Patient 117), which causes an aspartic acid to asparagine amino acid replacement, yielded two different transcripts: the full length wild-type and a 21 nucleotide shorter product. This effect was not observed in control fibroblast cell lines (Figure 1A). RT-PCR analysis of the c.1960G>A missense mutation (p.D654N; Patient 412) revealed partial skipping of exon 15 leading to a premature stop codon in exon 16. The aberrant fragment was absent in the control sample (Figure 1C). The c.6354C>T transition (Patients 58 and 245) does not lead to an amino acid exchange but RT-PCR analysis from fibroblast-derived RNA of Patient 58 and PAXgene blood-RNA of Patient 245 revealed skipping of the entire exon 51 (Figure 1B). The c.1981_1982insT insertion (Patient 262) is at the 21 nucleotide position of exon 16 and was originally interpreted as a frameshift mutation (p.C661L*fs*X20). Now, we can show the skipping of exon 16 (153 bp) based on cDNA analysis and subsequent sequencing. This transcript is 153 nucleotides shorter and presumably translated into shorter protein (p.D654_S704del). The remaining 66 heterozygous variants were not associated with aberrant splicing. In the majority of the samples (58) both alleles were expressed at comparable levels (Table 2).

To investigate the exonic splicing enhancer (ESE) status of these sequence variants we performed bioinformatic prediction by three commonly used web-based programs,

ESEfinder, RESCUE-ESE, together with PESX, and compared the results with the observed effect on RNA. As described in Table 2 and summarized in Table 4, of the 70 variants 48 were localized in ESE motifs identified by ESEfinder. The same data set was also analyzed by the RESCUE-ESE and PESX programs that identified 18 and 10, respectively. Compared these predictions with our *in vitro* results, out of the four splicing defect, only ESEfinder predicted one variant to lie in an ESE motif. The remaining three mutations were not predicted as putative ESE sequence. PESX and RESCUE-ESE were not able to predict these variants. These predictions were then applied to calculate the sensitivity, specificity, and accuracy of these three algorithms. As shown in Table 5, the highest sensitivity, which measures the proportion of true positives, was resulted by ESEfinder 3.0, while RESCUE-ESE and PESX did not identify true positives. The specificity, which measures the proportion of true negatives, proved to be much higher for RESCUE-ESE and PESX (73 and 93%), whereas ESEfinder was less reliable (29%). Further, we calculated the accuracy, which takes into account both true positive and true negative results. PESX (80%) and RESCUE-ESE (69%) showed the best accuracy, while ESEfinder (29%) proved to be less accurate for prediction of our samples.

In vitro and *in silico* Analyses of *FBN1* Intronic Variants

The consequence of 41 intronic substitutions was also investigated by cDNA analysis with primers spanning the affected exon, followed by qualitative sequencing and fragment length analysis. The GU-motif at the donor site was affected in nine cases, eight mutations were located in the AG-motif at the acceptor site, and the remaining 24 variants were outside of these canonical motifs (Table 3). All nine mutations abolished the splice donor site and revealed aberrant transcript processing, causing exon skipping, the activation of a novel cryptic donor site, intron retention, or even the combination of these events (e.g. Figure 1C and E). The eight splice acceptor site substitutions affected the mRNA splicing similarly (e.g. Figure 1D and F).

A significant fraction (24) of the intronic alterations lay outside of the canonical acceptor and donor sites. Seven out of these 24 alterations displayed aberrant splicing causing exon skipping, the activation of a novel cryptic donor or acceptor site, or a combination of these events. These variants; however, were mapped very close to the conserved canonical sequences. The only exception was the c.6314-15G>A variant, which led to an aberrant transcript lacking exon 51. In the remaining cases we did not observe aberrant splicing by cDNA analyses.

For *in silico* prediction of intronic variants we used five bioinformatic tools, the Human Splicing Finder version 2.4 (HSF) which also includes the MaxEntScan (MES) algorithm, the NNSPLICE version 0.9, the NetGene2 splice site prediction tools, and the splice site analysis function of the ESEfinder 3.0. Table 3 shows the differences between the wild-type and mutant scores provided by the above mentioned programs calculated either by us (NNSplice, NetGenee2, and ESEfinder 3.0) or the program itself (HSF and MES). The differences were calculated in percentage. Compared with our RNA-analyses we found excellent prediction for nine mutations in the canonical donor site (the first two nucleotides of the intron) for all algorithms. In two cases ESEfinder did not predict the splice site lost, while the other programs did. Out of the nine canonical acceptor variants six were predicted to lead splice site loss by all programs. While in two cases NNSplice, in one case NetGene2, and in three cases ESEfinder did not identify the splice site defects. For variants lying outside of the first two positions and last two positions of introns we could classify the predictions into two categories. Variants near to the splice sites or part of them were predicted correctly to cause splicing defect comparing with our experimental data. Mutations deeper in intron were not or just poorly identified to lie in a splice site motif.

The sensitivity, specificity, and accuracy of these programs for the previously investigated 41 intronic variants were calculated and summarized in Tables 6 and 7. We assessed the sensitivity and accuracy of HSF, MES, NNSplice, and NetGene2 at 100%, while ESEfinder seemed less sensitive and accurate (77-77%, respectively) at the highly conserved +1 and +2 donor site positions (Table 7A). Analysis of the acceptor site mutations at -1 and -2 positions proved to be more variable compared to the donor site predictions (Table 7B). The highest sensitivity and accuracy was achieved using HSF and MES (100%), whereas the other three programs achieved between 67% and 89% for sensitivity and accuracy, respectively (Table 7). The performance of the programs for intronic variants outside of the highly conserved acceptor and donor dinucleotides was very comparable, achieving between 63% and 100% sensitivity, specificity, and accuracy (Table 7C).

We further analyzed 14 intronic sequence variants without material for RNA-analysis *in silico* in order to predict their impact on splicing. Two out of 14 were part of the canonical donor site (Patients 498 and 213 in Table 3) and three were located on the acceptor site (Patients 30365 and 154). Predictions by the above used programs for these five variants resulted in the loss of the splice sites. From the remaining nine variants one affected the G nucleotide at the +5 position of intron (Patient 463 in Table 3). Four algorithms predicted the

loss of the splice site of this nucleotide change. The other eight variants were located at less or none conserved positions and they were not predicted as potential splice site.

In four of the exonic mutations splice defects were found. Therefore, we searched with the online tools described above for prediction. For only one of them, c.4192G>A (Patient 117), a new donor site (HSF: score of 84.59; NNSplice: 0.73; NetGene2 0.95) was predicted. The other one of the four, c.1960G>A (Patient 412), which affects the last nucleotide of the exon, all algorithms predicted a loss of the splice site. Finally, the c.6354C>T and the c.1981_1982insT variants were not predicted as potential splice site by any program.

In many cases, the agarose gel images displayed more then two expected fragments (Figure 1). Surprisingly, sequencing of the PCR fragments did not reveal a third template. Isolation and sequencing of these additional fragments indicated the presence of both, the wild-type and the mutant transcripts. This might be possible if heteroduplexes are formed. We selected several patients to test our hypothesis. We repeated the RT-PCR with one of the primers being marked with a fluorescent FAM label. Fragment analysis revealed the presence of two products each representing the expected size of the mutant and wild-type amplicon (Figure 2). Together these results strongly support an explanation of heteroduplex formations.

Discussion

In this study we investigated the consequences of 70 exonic and 55 intronic variants in the *FBN1* gene and used different bioinformatics tools to evaluate predictions with aberrant splicing detected by cDNA analysis.

Our results showed that from the 70 nucleotide substitutions one synonymous, two non-synonymous mutations, and one insertion were found to cause splicing defect. The c.6354C>T transition (Patients 58 and 245) which causes the skipping of the entire exon 51 was previously described in skin fibroblast strain, derived from patient with classical Marfan syndrome [Liu et al., 1997]. No *in silico* program exepting ESEfinder predicted altered splicing due to this variant. This program resulted in higher mutant score for the SC35 SR-protein which suggests that the protein might be involved in the regulation of splicing of this exon. It is interesting to note that the *FBN1* c.6339T>G (p.Y2113X) nonsense mutation was shown to cause the skipping of exon 51, as well [Dietz et al., 1993]. As this mutation is not fare away from the c.6354C>T transition we can speculate whether this region of exon 51 is a composite exonic regulatory element of splicing (CERES). These exonic regulatory elements contain overlapping enhancer and silencer motifs and can cause splicing abnormalities if mutation is appeared [Pagani et al., 2003]. However, the presence of a clear ESE-motif at this position is still possible because not all binding sites of important SR-proteins have been identified, and further, the known binding motifs are relatively poorly defined.

The c.4192G>A missense mutation does not seem to be an ESE-sequence. It is more likely a cryptic splice site activated in transcripts bearing the mutation. Splice site prediction tools predicted a high strength of the putative splice site, even as high as the natural site. The c.1960G>A missense variant affects the last nucleotide of exon 15. Nucleotides at this position are part of the 5` consensus sequence and important for the interaction with the U1 snRNA. Since the G is highly conserved at this position (92% for *FBN1*), an alteration is expected to affect splicing. The third exonic mutation, c.1981_82insT, also caused aberrant transcript processing. The potential existence of an ESE could not be confirmed. In normal *FBN1* transcripts PESX program predicts a splice silencer motif from 1980 to 1987. The insertion mutation (c.1981_82insT) destroys this silencer motif. Whether this has consequences for splicing can not be answered at this time.

The remaining exonic variants were not associated with aberrant splicing. These data suggest that the majority of *FBN1* missense mutations we tested most likely impair the biological nature of the protein rather than the splicing pattern. On the other hand, these

variants can also affect mRNA stability. Recent studies indicated that both synonymous and non synonymous SNPs can disturb the secondary structure of the mRNA, as showed with the silent mutation in the dopamine receptor D2 gene (*DRD2*) or in the multidrug resistance 1 gene (*MDR1*). These studies illustrated that polymorphisms in *DRD2* and *MDR1* genes affect mRNA stability and thereby reduce mRNA and protein expression [Duan et al., 2003; Wang et al., 2005]. Moreover, silent mutations can even affect the timing of cotranslational folding and this way altering the conformation. Recently, a study reported three previously known synonymous mutations in the *MDR1* gene which alter the encoded transmembrane protein, the P-glycoprotein (P-gp) activity [Kimchi-Safarty et al., 2007]. They found that the P-gp inhibitors cyclosporine A and verapamil were less effective against the proteins contain the combinations of the SNPs. The authors argued that the codon usage for the SNP (frequent versus infrequent) may influence the translation rate and thereby affects the folding of the protein. We also analyzed our nine silent mutations on the same way, in order to see whether there is a difference in the codon usage. Eight out of the nine variants showed frequent codon change to rare and one rare to frequent (Table 6). We can speculate that these variants can also be pathogenic by influencing the translational rate which in turn affects the protein folding.

In 6 cases (Patients 125, 319, 19, 263, 26, and T29668) we did not found splicing defect in spite of not detectable or reduced mutant allele amount. These results suggest that those mutations affect rather the expression level of mRNA providing allelic variation. Allelic variation in expression between two alleles is common in human genes. Recently, number of studies reported numerous genes and showed that 20-50% of human genes are not equally expressed [Yan et al., 2002; Kim et al., 2008]. Allelic variation is may be the basis of the human variation [Lo et al., 2003]. Therefore, measurement of the allelic expression is recommended for RNA studies. All nucleotide changes in the canonical donor (GT) and acceptor (AG) sites resulted in aberrant splicing. The splicing process was affected by different ways: exon skipping, creation of cryptic splice site(s) by partial deletion of the exonic or insertion of intronic sequences. Additionally, combination of these events was also occured. Nucleotide substitutions near the canonical splice donor and acceptor sites displayed similar consequences as substitutions in the canonical dinucleotides. While variants, lying deeper in introns did not show aberrant splicing, suggesting those alterations as common polymorphic variants. Our results regarding the effect of variants lying at or near to the consensus splice sites are consistent with previous findings in which *BRCA1*, *BRCA2*, and *RB1* genes were investigated [Auclair et al., 2006; Bonatti et al., 2006; Lastella et al., 2006;

Bonnet et al., 2008; Houdayer et al., 2008; Arnold et al., 2009]. It is important to note that in many of these studies the investigated missense or silent mutations leading to splicing defects lie very close to or in the exonic part of the consensus sequence. This raises the question whether those nucleotides disturb an ESE or the splice site. It is more likely the latter case because the U1 snRNA:pre-mRNA interaction include the nucleotide position -3 to +8 at the donor site [Hartmann et al., 2008].

Interestingly, the c.4337_4339delATA (Patient 7) small deletion which affects the first three nucleotides of exon 35, and the c.5066A>G (Patient 395) missense mutation at the first position of exon 41 did not resulted aberrant transcript processing. This was also supported with the fact that quantitative sequence analysis resulted in comparable wild-type and mutant allele expression. In order to see, whether there is a difference between sequence conservation and nucleotide usage of *FBN1* and other genes, we compared all *FBN1* splice sites (-3, +8 position for the donor site and -14, +2 position for the acceptor site) with annotated eukaryotic splice sites (Supplementary Figure 1) [Hartmann et al., 2008]. We found that at the +1 exon position of the *FBN1* acceptor site the A appears to be more frequenly used as in other eukaryotic genes where all the nucleotides can occur. This observation can explore why c.4337_4339delATA and c.5066A>G did not revealed splicing defect. It seems that the new nucleotides are also tolerated at those positions.

Regarding the highly conserved splice consensus sites we could see that the *FBN1* splice sites also possess GT-AG consensus sequences with one exception. At the 5` end of intron 62 a GC dinucleotides can be found instead of GT. This GC site is a most common class of nonconsensus splice sites, which is assumed to recognize by the standard (U2-dependent) spliceosome [Wu and Krainer, 1999; Thanaraj and Clark, 2001]. It should be mentioned, that ~62% of the GC-AG introns are alternative introns and a large part of those are isoforms of GT-AG introns [Thanaraj and Clark, 2001]. This has raised the question whether exon 62 of the *FBN1* gene is an alternatively spliced exon. We found that exon 62 included in transcript using both blood and fibroblasts-RNA for analysis. However, we can not exclude that in other tissues exon 62 or a part of those retained, as well.

This study further allowed us the comparison of splicing patterns in PAXgene blood-RNA (Patient 245) with those in fibroblast cell lines (Patient 58). We found the same splicing pattern in both tissues (Figure 1B). This observation can very helpful on the analysis of the effect of exonic and intronic variants found at genomic DNA level. PAXgene blood is relatively easy available and no cell culture is needed, which makes transcript analyses difficult and time consuming. Another non-invasive alternative can be the RNA extraction

from saliva. Recently, several companies brought such systems on the market, which allow total RNA extraction from saliva (e.g. Qiagen RNeasy Protect Saliva Mini Kit, Oragene•RNA for Expression Analysis Self-Collection Kit). We tested the Oragene•RNA product in our laboratory and were able to amplify *FBN1* transcripts from saliva (data not shown). This non-invasive and easy-to-use kit could provide a good alternative for fast and reliable transcript analysis for testing the pathogenic effect of variants. However, further analyses are needed to clarify whether cDNA-analysis using RNA from saliva is representative for extracellular matrix.

Taken together, this is the first systematic *FBN1* mRNA screening for aberrant transcript processing in a total of 125 unrelated MFS patient with different exonic and intronic nucleotide substitutions. In this paper we provided evidence for aberrant splicing caused by missense and silent mutations and by intronic variants differed from the consensus splice sites. We have also shown that the three exonic splicing enhancer programs were not very reliable in our case. While the applied *in silico* splice site prediction programs were sensitive enough to predict the loss of splice sites; however, the use of ectopic splice sites was not well predicted. It is noteworthy that neither the used *in silico* programs are validated for diagnostic use. In their present state they are not able to replace *in vitro* studies. They can be used rather to prioritize variants for *in vitro* analysis and confirmation of disturbed splicing, as it was suggested recently [Arnold et al., 2009]. Consequently, RT-PCR analysis is required to assess pathogenicity and for correct interpretation of variants on pre-mRNA splicing.

Acknowledgement

We are grateful to the patients for participating in this study and the referring physicians (Klaus Ammann, Deborah Bartholdi, Sophie Dahoun, Alexandre Dayer, Nursel Elcioglu, Claudio Fabris, Bruno Frischkopf, Monica Gersbach-Forrer, Gerber Glanzmann, Oswald Hasselmann, Karl Heinimann, Peter Hoessly, Daniela Kaiser, Thomas Lüscher, Hansjakob Müller, Kirsten Rasmussen, Albert Schinzel, Roger Simon, Barbara Utermann, and Hari Zvizdic) for help with obtaining appropriate samples and/or information about the clinical phenotypes of patients. We thank Angelika Schwarze for cell cultures; Melanie Maudrich for initial *FBN1* mutation analysis; Philippe Reuge for technical support; Barbara Kloeckener, John Neidhardt, Fabian Schmid, Caroline Henggeler, Regina Perez, and other members of the Institute of Medical Molecular Genetics, University of Zurich, for constructive discussions.

We also would like to thank Regina Perez for reading our manuscript. We would like to thank Adrian R. Krainer and Luca Cartegni for helpful information and comment regarding ESEfinder. This work was supported by grants from the Foundation for Research at the Medical Faculty and Research Funding of the University of Zurich (to I.M. and G.M.), Swiss Heart Foundation (to G.M.), Swiss National Science Foundation (NF 3200B0-109370/1 to B.S.; NF 3100A0-120504 to G.M.), and the Wolfermann-Nägeli-Stiftung Zurich (to B.S.).

References

Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, Woods MO, Hopper JL, Jenkins MA, Brown MA, Tavtigian SV, Goldgar DE, Young JP, Spurdle AB. 2009. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. Hum Mutat 30(5):757-770.

Auclair J, Busine MP, Navarro C, Ruano E, Montmain G, Desseigne F, Saurin JC, Lasset C, Bonadona V, Giraud S, Puisieux A, Wang, Q. 2006. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. Hum Mutat 27(2):145-154.

Baralle D, Lucassen A, Buratti E. 2009. Missed threads. The impact of pre-mRNA splicing defects on clinical practice. EMBO Rep 10(8):810-816.

Baumgartner C, Mátyás G, Steinmann B, Baumgartner D. 2005. Marfan syndrome - a diagnostic challenge caused by phenotypic and genetic heterogeneity. Methods Inf Med 44(4):487-497.

Bonatti F, Pepe C, Tancredi M, Lombardi G, Aretini P, Sensi E, Falaschi E, Cipollini G, Bevilacqua G, Caligo MA. 2006. RNA-based analysis of BRCA1 and BRCA2 gene alterations. Cancer Genet Cytogenet 170(2):93-101.

Bonnet C, Krieger S, Vezain M, Rousselin A, Tournier I, Martins A, Berthet P, Chevrier A, Dugast C, Layet V, Rossi A, Lidereau R, Frebourg T, Hardouin A, Tosi M. 2008. Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription-PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. J Med Genet 45(7):438-446.

Booms P, Cisler J, Mathews KR, Godfrey M, Tiecke F, Kaufmann UC, Vetter U, Hagemeier C, Robinson PN. 1999. Novel exon skipping mutation in the fibrillin-1 gene: two 'hot spots' for the neonatal Marfan syndrome. Clin Genet 55(2):110-117.

Brunak S, Engelbrecht J, Knudsen S. 1991. Prediction of human mRNA donor and acceptor sites from the DNA sequence. J Mol Biol 220(1):49-65.

Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3(4):285-298.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 31(13):3568-3571.

Christodoulou J, Petrova-Benedict R, Robinson BH, Jay V, Clarke JT. 1993. An unusual patient with the neonatal Marfan phenotype and mitochondrial complex I deficiency. Eur J Pediatr 152(5):428-432.

Collod-Beroud G, Beroud C, Ades L, Black C, Boxer M, Brock DJ, Holman KJ, de Paepe A, Francke U, Grau U, Hayward C, Klein HG, Liu W, Nuytinck L, Peltonen L, Alvarez Perez AB, Rantamäki T, Junien C, Boileau C. 1998. Marfan Database (third edition): new mutations and new routines for the software. Nucleic Acids Res 26(1):229-233.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14(6):1188-1190.

De Paepe A, Devereux RB, Dietz HC, Hennekam RC, Pyeritz RE. 1996. Revised diagnostic criteria for the Marfan syndrome. Am J Med Genet 62(4):417-426.

Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 37(9):e67.

Dietz HC, Cutting GR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, Puffenberger EG, Hamosh A, Nanthakumar EJ, Curristin SM, Stetten G, Meyers DA, Francomano CA. 1991. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. Nature 352(6333):337-339.

Dietz HC, Valle D, Francomano CA, Kendzior RJ, Jr., Pyeritz RE, Cutting GR. 1993. The skipping of constitutive exons in vivo induced by nonsense mutations. Science 259(5095):680-683.

Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J, Gejman PV. 2003. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. Hum Mol Genet 12(3):205-216.

Fairbrother WG, Yeo GW, Yeh R, Goldstein P, Mawson M, Sharp PA, Burge CB. 2004. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. Nucleic Acids Res 32(Web Server issue):W187-190.

Godfrey M, Vandemark N, Wang M, Velinov M, Wargowski D, Tsipouras P, Han J, Becker J, Robertson W, Droste S, Rao VH. 1993. Prenatal diagnosis and a donor splice site mutation in fibrillin in a family with Marfan syndrome. Am J Hum Genet 53(2):472-480.

Halliday DJ, Hutchinson S, Lonie L, Hurst JA, Firth H, Handford PA, Wordsworth P. 2002. Twelve novel FBN1 mutations in Marfan syndrome and Marfan related phenotypes test the feasibility of FBN1 mutation testing in clinical practice. J Med Genet 39(8):589-593.

Hartmann L, Theiss S, Niederacher D, Schaal H. 2008. Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? Front Biosci 13:3252-3272.

Houdayer C, Dehainault C, Mattler C, Michaux D, Caux-Moncoutier V, Pages-Berhouet S, d'Enghien CD, Lauge A, Castera L, Gauthier-Villars M, Stoppa-Lyonnet D. 2008. Evaluation of in silico splice tools for decision-making in molecular diagnosis. Hum Mutat 29(7):975-982.

Howarth R, Yearwood C, Harvey JF. 2007. Application of dHPLC for mutation detection of the fibrillin-1 gene for the diagnosis of Marfan syndrome in a National Health Service Laboratory. Genet Test 11(2):146-152.

Kainulainen K, Karttunen L, Puhakka L, Sakai L, Peltonen L. 1994. Mutations in the fibrillin gene responsible for dominant ectopia lentis and neonatal Marfan syndrome. Nat Genet 6(1):64-69.

Katzke S, Booms P, Tiecke F, Palz M, Pletschacher A, Turkmen S, Neumann LM, Pregla R, Leitner C, Schramm C, Lorenz P, Hagemeier C, Fuchs J, Skovby F, Rosenberg T, Robinson PN. 2002. TGGE screening of the entire FBN1 coding sequence in 126 individuals with marfan syndrome and related fibrillinopathies. Hum Mutat 20(3):197-208.

Kim YS, Choi YB, Lee JH, Yang SH, Cho JH, Shin CH, Lee SD, Paik MK, Hong KM. 2008. Predisposition of genetic disease by modestly decreased expression of GCH1 mutant allele. Exp Mol Med 40(3):271-275.

Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. 2007. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science 315(5811):525-528.

Lastella P, Surdo NC, Resta N, Guanti G, Stella A. 2006. In silico and in vivo splicing analysis of MLH1 and MSH2 missense mutations shows exon- and tissue-specific effects. BMC Genomics 7:243.

Liu W, Qian C, Francke U. 1997. Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. Nat Genet 16(4):328-329.

Lo HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP. 2003. Allelic variation in gene expression is common in the human genome. Genome Res 13(8):1855-1862.

Loeys B, Nuytinck L, Delvaux I, De Bie S, De Paepe A. 2001. Genotype and phenotype analysis of 171 patients referred for molecular study of the fibrillin-1 gene FBN1 because of suspected Marfan syndrome. Arch Intern Med 161(20):2447-2454.

Magyar I, Colman D, Arnold E, Baumgartner D, Bottani A, Fokstuen S, Addor MC, Berger W, Carrel T, Steinmann B, Mátyás G. 2009. Quantitative sequence analysis of FBN1 premature termination codons provides evidence for incomplete NMD in leukocytes. Hum Mutat 30(9):1355-1364.

Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. Trends Genet 24(3):133-141.

Mátyás G, De Paepe A, Halliday D, Boileau C, Pals G, Steinmann B. 2002. Evaluation and application of denaturing HPLC for mutation detection in Marfan syndrome: Identification of 20 novel mutations and two novel polymorphisms in the FBN1 gene. Hum Mutat 19(4):443-456.

Pagani F, Stuani C, Tzetis M, Kanavakis E, Efthymiadou A, Doudounakis S, Casals T, Baralle FE. 2003. New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. Hum Mol Genet 12(10):1111-1120.

Perez AB, Pereira LV, Brunoni D, Zatz M, Passos-Bueno MR. 1999. Identification of 8 new mutations in Brazilian families with Marfan syndrome. Mutations in brief no. 211. Hum Mutat 13(1):84.

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. J Comput Biol 4(3):311-323.

Rommel K, Karck M, Haverich A, Schmidtke J, Arslan-Kirchner M. 2002. Mutation screening of the fibrillin-1 (FBN1) gene in 76 unrelated patients with Marfan syndrome or Marfanoid features leads to the identification of 11 novel and three previously reported mutations. Hum Mutat 20(5):406-407.

Rommel K, Karck M, Haverich A, von Kodolitsch Y, Rybczynski M, Muller G, Singh KK, Schmidtke J, Arslan-Kirchner M. 2005. Identification of 29 novel and nine recurrent fibrillin-1 (FBN1) mutations and genotype-phenotype correlations in 76 patients with Marfan syndrome. Hum Mutat 26(6):529-539.

Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 18(20):6097-6100.

Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM. 2008. Prediction and assessment of splicing alterations: implications for clinical testing. Hum Mutat 29(11):1304-1313.

Stahl-Hallengren C, Ukkonen T, Kainulainen K, Kristofersson U, Saxne T, Tornqvist K, Peltonen L. 1994. An extra cysteine in one of the non-calcium-binding epidermal growth factor-like motifs of the FBN1 polypeptide is connected to a novel variant of Marfan syndrome. J Clin Invest 94(2):709-713.

Stheneur C, Collod-Beroud G, Faivre L, Buyck JF, Gouya L, Le Parc JM, Moura B, Muti C, Grandchamp B, Sultan G, Claustres M, Aegerter P, Chevallier B, Jondeau G, Boileau C. 2009. Identification of the minimal combination of clinical features in probands for efficient mutation detection in the FBN1 gene. Eur J Hum Genet 17(9):1121-1128.

Thanaraj TA, Clark F. 2001. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. Nucleic Acids Res 29(12):2581-2593.

Thusberg J, Vihinen M. 2009. Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. Hum Mutat 30(5):703-714.

Tiecke F, Katzke S, Booms P, Robinson PN, Neumann L, Godfrey M, Mathews KR, Scheuner M, Hinkel GK, Brenner RE, Hövels-Gürich HH, Hagemeier C, Fuchs J, Skovby F, Rosenberg T. 2001. Classic, atypically severe and neonatal Marfan syndrome: twelve mutations and genotype-phenotype correlations in FBN1 exons 24-40. Eur J Hum Genet 9(1):13-21.

Tucker T, Marra M, Friedman JM. 2009. Massively parallel sequencing: the next big thing in genetic medicine. Am J Hum Genet 85(2):142-154.

Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. 2002. Allelic variation in human gene expression. Science 297(5584):1143.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 11(2-3):377-394.

Yuan B, Thomas JP, von Kodolitsch Y, Pyeritz RE. 1999. Comparison of heteroduplex analysis, direct sequencing, and enzyme mismatch cleavage for detecting mutations in a large gene, FBN1. Hum Mutat 14(5):440-446.

Wang D, Johnson AD, Papp AC, Kroetz DL, Sadee W. 2005. Multidrug resistance polypeptide 1 (MDR1, ABCB1) variant 3435C>T affects mRNA stability. Pharmacogenet Genomics 15(10):693-704.

Wu Q, Krainer AR. 1999. AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. Mol Cell Biol 19(5):3225-3236.

Zhang XH, Chasin LA. 2004. Computational definition of sequence motifs governing constitutive exon splicing. Genes Dev 18(11):1241-1250.

Zhang XH, Kangsamaksin T, Chao MS, Banerjee JK, Chasin LA. 2005. Exon inclusion is dependent on predictable exonic splicing enhancers. Mol Cell Biol 25(16):7323-7332.

TABLE 1. *FBN1* Sequence Variants with Clinical Features

Patient						Sequence variant[c]				
#[a]	Year of birth	Phenotype[b]				Location		Nucleotide change	AA change	Reference
		SS	OS	CS	FH	UTR/Exon/Intron (size in nts)	Position of the change			
5`UTR and exonic mutations										
125	1968	-	-	+	-	5`UTR (181)	46	c.-136G>C	-	This study
159	1945	-	-	+	-	Exon 1 (164)	61	c.61A>G	p.T21A	This study
248	1968	(+)	+	+	-	Exon 2 (83)	75	c.239G>A	p.C80Y	[Howarth et al., 2007]
91	1993	(+)	+	+	-	Exon 3 (99)	19	c.266G>A	p.C89Y	This study
259	1981	(+)	+	+	-	Exon 4 (96)	18	c.364C>T	p.R122C	[Stahl-Hallengren et al., 1994]
265	1997	(+)	+	-	-	Exon 4 (96)	18	c.364C>T[§]	p.R122C[§]	[Stahl-Hallengren et al., 1994]
293	1964	(+)	-	(+)	-	Exon 6 (198)	75	c.613A>G	p.T205A	This study
T473	1962	?	?	+	?	Exon 8 (126)	124	c.986T>C	p.I329T	[Stheneur et al., 2009]
14	1998	+	+	+	?	Exon 9 (159)	142	c.1130G>A	p.C377Y	This study
212	1958	(+)	?	(+)	+	Exon 10 (180)	30	c.1177A>G	p.M393V	This study
319	1968	(+)	+	+	-	Exon 15 (123)	13	c.1850G>A	p.C617Y	This study
319B	2006	?	?	?	?	Exon 15 (123)	13	c.1850G>A[†]	p.C617Y[†]	This study
254	1951	(+)	(+)	+	-	Exon 15 (123)	39	c.1876G>A	p.G626R	This study
454	1982	+	+	+	-	Exon 15 (123)	46	c.1883G>T	p.C628F	This study
412	1972	(+)	?	+	+	Exon 15 (123)	123	c.1960G>A	p.D654N	[Halliday et al., 2002]
262	1995	(+)	-	+	-	Exon 16 (153)	21	c.1981_1982insT	p.C661L*fs*X20	[Magyar et al., 2009]
T472	1952	((+))	-	+	-	Exon 23 (126)	49	c.2777delG	p.C926L*fs*X16	This study
SV	1993	?	-	(+)	+	Exon 23 (126)	76	c.2804G>A	p.C935Y	This study
SS	1972	?	+	(+)	-	Exon 23 (126)	76	c.2804G>A[†]	p.C935Y[†]	This study
218	1953	-	(?)	+	-	Exon 24 (228)	102	c.2956G>A	p.A986T	This study
270	2000	+	-	(+)	-	Exon 25 (126)	34	c.3116G>A	p.C1039Y	[Liu et al., 1997]
321	1978	(+)	+	+	+	Exon 25 (126)	67	c.3149G>C	p.S1050T	This study
WR1	1979	?	?	?	?	Exon 26 (129)	12	c.3220T>C	p.C1074R	[Kainulainen et al., 1994]
289	1997	(+)	+	(+)	+	Exon 26 (129)	42	c.3250G>C	p.G1084R	This study
339	1979	(+)	(+)	(+)	?	Exon 28 (126)	40	c.3503A>G	p.N1168S	This study
252	1969	+	+	+	?	Exon 28 (126)	82	c.3545G>A	p.C1182Y	This study
11	1926	?	?	?	?	Exon 30 (126)	39	c.3751G>T	p.G1251C	This study
71	1987	+	+	+	-	Exon 30 (126)	49	c.3761G>A	p.C1254Y	[Stheneur et al., 2009]

380	2004	+	+`	+	?	Exon 30 (126)	120	c.3832T>A	p.C1278S	This study
112	1990	+	+	+	-	Exon 31 (126)	48	c.3886T>C	p.C1296R	This study
381	2000	-	+	+	-	Exon 31 (126)	91	c.3929G>T	p.G1310V	[Howarth et al., 2007]
187	1942	?	?	?	?	Exon 32 (123)	10	c.3974A>C	p.E1325A	This study
107	1980	?	+	+	-	Exon 32 (123)	61	c.4025C>T	p.T1342I	This study
89	1995	((+))	+	?	+	Exon 33 (123)	4	c.4091_4095delinsCATTG	p.L1364_D1365delinsPL	This study
19	1995	+/-	+	+	-	Exon 33 (123)	34	c.4121G>A	p.C1374Y	[Stheneur et al., 2009]
263	2000	+	+	+	-	Exon 33 (123)	56	c.4143_4145delGAA	p.K1381del	This study
114	1958	+	-	+	?	Exon 33 (123)	79	c.4166G>A	p.C1389Y	This study
117	1983	?	+	+	+	Exon 33 (123)	105	c.4192G>A	p.D1398N	This study
38	1962	-	-	+	-	Exon 34 (126)	7	c.4217A>G	p.D1406G	[Tiecke et al., 2001]
382	2003	+	+	(+)	-	Exon 34 (126)	12	c.4222T>C	p.C1408R	[Stheneur et al., 2009]
261	1975	(+)	-	(+)	?	Exon 34 (126)	60	c.4270C>G	p.P1424A	[Collod-Beroud et al., 1998]
163	1943	-	-	+	-	Exon 34 (126)	96	c.4306G>T	p.V1436L	This study
7	1978	+	+	+	+	Exon 35 (123)	1	c.4337_4339delATA	p.D1446_I1447delinsV	[Baumgartner et al., 2005]
KB	1957	?	?	?	?	Exon 35 (123)	73	c.4409G>A	p.C1470Y	[Mátyás et al., 2002]
4B	1957	+	+	+	-	Exon 35 (123)	118	c.4454G>A	p.C1485Y	[Howarth et al., 2007]
39	1970	+	-	(+)	+	Exon 36 (123)	52	c.4511A>G	p.N1504S	This study
78	1950	(+)	+	+	+	Exon 36 (123)	78	c.4537T>C	p.C1513R	[Kainulainen et al., 1994]
95	1941	?	+	+	+	Exon 37 (165)	6	c.4588C>T	p.R1530C	[Loeys et al., 2001]
395	1986	(+)	+	-	-	Exon 41 (159)	1	c.5066A>G	p.D1689G	This study
321	1978	(+)	+	+	+	Exon 41 (159)	19	c.5084G>A	p.C1695Y	This study
121	1942	-	+	+	-	Exon 45 (126)	9	c.5554G>A	p.E1852K	This study
73	1992	+	+	+	-	Exon 46 (117)	55	c.5726T>C	p.I1909T	[Loeys et al., 2001]
260	1992	-	-	?	+	Exon 46 (117)	105	c.5776_5778delAAT	p.N1926del	This study
292	1998	+	+	(?/-)	-	Exon 49 (126)	32	c.6069_6080del	p.C2024_G2027del	This study
459	1957	(?)	(+)	+	(?)	Exon 51 (66)	19	c.6332G>T	p.C2111F	This study
26	1959	+	-	+	+	Exon 53 (120)	81	c.6577G>A	p.E2193K	This study
T29668	1949	(+)	-	+	+	Exon 53 (120)	81	c.6577G>A[§]	p.E2193K[§]	This study
493	1984	?	?	?	?	Exon 53 (120)	114	c.6610T>C	p.C2204R	This study
150	1988	+	-	(+)	-	Exon 54 (123)	12	c.6628T>C	p.C2210R	This study
9	1969	+/-	-	+/-	+	Exon 63 (232)	9	c.7828G>A	p.E2610K	[Liu et al., 1997]
17	1971	?	?	?	+	Exon 63 (232)	33	c.7852G>A	p.G2618R	[Mátyás et al., 2002]
173	1987	+	-	-	+	Exon 65 (390)	380	c.8606T>C	p.L2869S	This study
Exonic polymorphisms										
T30783	1952	(+)	-	+	-	Exon 5 (96)	68	c.510C>T	p.Y170Y	This study
270	2000	+	-	(+)	-	Exon 6 (198)	101	c.639C>T	p.V213V	This study
295B	1971	?	?	?	?	Exon 13 (126)	80	c.1668G>A	p.V556V	This study

T31083	1996	(+)	-	-	-	Exon 15 (123)	8	c.1845C>T	p.N615N	This study
345	1993	+	-	+	-	Exon 24 (228)	41	c.2895G>A	p.E965E	This study
373	1965	(+)	-	+	+	Exon 26 (129)	86	c.3294C>T	p.D1098D	[Yuan et al., 1999]
33	1960	-	-	(+)	?	Exon 35 (123)	92	c.4428C>T	p.Y1476Y	This study
58	1984	?	?	?	?	Exon 51 (66)	41	c.6354C>T	p.I2118I	[Liu et al., 1997]
245	1951	(+)	+	+	+	Exon 51 (66)	41	c.6354C>T[§]	p.I2118I[§]	[Liu et al., 1997]
T473	1962	?	?	+	?	Exon 52 (117)	14	c.6393C>T	p.C2131C	This study
150	1988	+	-	(+)	-	Exon 54 (123)	65	c.6681A>C	p.S2227S	[Perez et al., 1999]
150B	1951	-	-	-	-	Exon 54 (123)	65	c.6681A>C[†]	p.S2227S[†]	[Perez et al., 1999]
319	1968	(+)	+	+	-	Exon 58 (126)	5	c.7209C>A	p.I2403I	This study
30263	1948	?	(+)	+	(+)	Exon 65 (126)	96	c.8322G>A	p.K2774K	This study
Intronic variants										
498	1982	+	-	+	?	Intron 1 (31513)	+1	c.164+1G>T	-	This study
31326	1971	(+)	-	+	?	Intron 1 (31513)	-86	c.165-86C>A	-	This study
31326	1971	(+)	-	+	?	Intron 2 (2183)	-44	c.248-44G>A	-	This study
284	1949	?	?	+	(+)	Intron 3 (10493)	+26	c.346+26A>G	-	This study
506	1999	?	?	?	?	Intron 3 (10493)	-73	c.347-73A>G	-	This study
369	1950	?	?	?	?	Intron 3 (10493)	-73	c.347-73A>G[§]	-	This study
305	1965	+	(+)	+	+	Intron 5 (58474)	-138	c.539-138C>T	-	This study
284	1949	?	?	+	(+)	Intron 6 (3405)	-26	c.737-26delT	-	[Christodoulou et al., 1993]
30783	1952	(+)	-	+	-	Intron 6 (3405)	-32	c.737-32T>G	-	This study
459	1957	(?)	(+)	+	(?)	Intron 7 (7824)	-10	c.863-10dupT	-	This study
29276	1953	-	-	+	-	Intron 7 (7824)	-121	c.863-121T>G	-	This study
287	1965	+	(+)	+	+	Intron 11 (1718)	+5	c.1468+5G>C	-	This study
28	1972	+	+/-	+	(+)	Intron 12 (3379)	-1-+2	c.1589-1_1591delins11	-	This study
24	1961	+	-	+	-	Intron 13 (1339)	+1	c.1714+1G>T	-	[Baumgartner et al., 2005]
75	1933	?	?	?	?	Intron 13 (1339)	+37	c.1714+37A>G	-	This study
30365	2003	?	?	?	?	Intron 13 (1339)	-2	c.1715-2A>G	-	This study
T30783	1952	(+)	-	+	-	Intron 14 (3434)	+16	c.1837+16C>T	-	This study
3	1983	+	+	+	+	Intron 14 (3434)	-2	c.1838-2A>G	-	[Baumgartner et al., 2005]
69B	1957	+	+	+	-	Intron 17 (1593)	-1	c.2168-1G>C	-	This study
352	1996	?	?	+	?	Intron 21 (868)	-70	c.2678-70C>G	-	This study
54	1953	+	+	+	+	Intron 22 (1617)	+1	c.2728+1G>T	-	This study
84	1989	?	?	?	?	Intron 27 (111)	+1	c.3463+1G>C	-	This study
T31083	1996	(+)	-	-	-	Intron 28 (1578)	-1	c.3590-1G>A	-	This study
146	2002	+	?	+	?	Intron 30 (2037)	+1	c.3838+1G>A	-	[Booms et al., 1999]
258	1939	+	-	(+)	(+)	Intron 31 (7004)	+61	c.3964+61A>G	-	This study
154	1985	+	(?)	+	+	Intron 34 (1794)	-1	c.4337-1G>A	-	This study

175	1984	(+)	-	(+)	?	Intron 34 (1794)	-29	c.4337-29T>A	-	This study
31174	1982	+	-	+	-	Intron 36 (309)	-5	c.4583-5A>G	-	This study
30964	1965	(+/-)	-	+	-	Intron 37 (2079)	+82	c.4747+82T>C	-	This study
65	1926	-	-	+	-	Intron 41 (2764)	-13	c.5225-13T>G	-	This study
249	1991	+	?	+	?	Intron 41 (2764)	-19	c.5225-19C>T	-	This study
277	1981	((+))	-	+	-	Intron 42 (3483)	+1	c.5296+1G>A	-	This study
289	1997	(+)	+	(+)	+	Intron 42 (3483)	+30	c.5296+30C>A	-	This study
6	1981	+	+	+	-	Intron 45 (1945)	+3	c.5671+3_5671+4insCC	-	This study
358	1982	+	-	+	?	Intron 46 (1201)	+10	c.5788+10C>A	-	This study
334	2004	?	?	?	?	Intron 46 (1201)	-1	c.5789-1G>C	-	This study
513	1991	+	+	(+)	?	Intron 48 (2694)	+2	c.6037+2T>C	-	This study
459	1957	(?)	(+)	+	(?)	Intron 48 (2694)	+40	c.6037+40_50del11	-	This study
410	1990	(+)	+	+	?	Intron 49 (3803)	-1	c.6164-1G>A	-	[Rommel et al., 2005]
514	1980	?	+	?	+	Intron 50 (380)	+53	c.6313+53A>G	-	This study
52	1970	+	+	?	?	Intron 50 (380)	-15	c.6314-15G>A	-	This study
463	1986	(+)	(+)	(+)	-	Intron 52 (2247)	+5	c.6496+5G>C	-	This study
94	1956	?	-	+	+	Intron 53 (1605)	+1	c.6616+1G>C	-	[Godfrey et al., 1993]
29203	1978	?	?	?	?	Intron 55 (2199)	-1	c.6872-1G>T	-	This study
394	2004	(+)	+	-	?	Intron 56 (572)	+1	c.6997+1G>T	-	This study
188	1974	?	?	?	?	Intron 56 (572)	+4-+13	c.6997+4_6997+13delinsT	-	This study
82	1965	?	+	+	?	Intron 56 (572)	-1	c.6998-1G>C	-	This study
383	1971	-	+	+	?	Intron 57 (1702)	+7	c.7204+7C>G	-	[Katzke et al., 2002]
281B	1977	?	?	?	?	Intron 57 (1702)	+54	c.7204+54dupT	-	This study
152	1958	(+)	-	+	+	Intron 58 (247)	+3-+8	c.7330+3_7330+8delins15	-	This study
T479	1965	(+)	-	+	-	Intron 59 (3300)	+1	c.7453+1G>C	-	This study
213	1966	?	?	?	+	Intron 59 (3300)	+1	c.7453+1G>A	-	This study
88	1936	+	+	+	+	Intron 61 (751)	+5	c.7699+5delG	-	This study
88B	1966	+	-	+	+	Intron 61 (751)	+5	c.7699+5delG†	-	This study
30205	1973	?	?	?	?	Intron 61 (751)	+5	c.7699+5G>A	-	This study
231	1976	+	+	-	+	Intron 62 (4919)	-2	c.7820-2A>T	-	This study
371	1988	+	+	+	-	Intron 63 (2792)	-23	c.8052-23T>C	-	[Rommel et al., 2002]

[a] Coded designations (two-letter code or numbers).

[b] SS, skeletal system; OS, ocular system; CS, cardiovascular system; FH, family history; +, affected; (+), slightly affected; -, not affected; ?, lack of information. Classification is based on diagnostic criteria for Marfan syndrome [De Paepe et al., 1996].

[c] Sequence variants are described in relation to the translation initiation site of the mRNA reference sequence NM_000138.3. Mutation numbering is after HGVS and journal standards, with +1 as the A of the ATG initiation codon. Recurrent mutations in related (†) and unrelated (§) patients are marked.

TABLE 2. *In vitro* and *in silico* Analyses of *FBN1* Exonic Sequence Variants

#[a]	Sequence variant[b]		RNA[c] source	*In silico* analysis												Aberrant splicing by *in vitro* cDNA analysis	Relative mutant allele expression (%)[g]
	Nucleotide change	AA change		ESEfinder[d]								RESCUE-ESE[e]		PESX[f]			
				SF2/ASF		SC35		SRp40		SRp55							
				Wt score	Mut score	Wt score	Mut score	Wt score	Mut score	Wt score	Mut score	Wt motif	Mut motif	Wt motif	Mut motif		
Putative 5' UTR and exonic mutations																	
125	c.-136G>C	-	B	3.71	3.96	-	-	-	-	-	-	2	-	3	-	Not detected	ND
159	c.61A>G	p.T21A	F, B	3.64	-	2.67	3.12	5.35	2.8	-	-	-	-	-	-	Not detected	~
248	c.239G>A	p.C80Y	B	-	-	-	-	-	-	-	-	-	-	-	-	Not detected	~60
91	c.266G>A	p.C89Y	F	-	-	3.26	2.81	2.95	3.11	-	-	-	-	1	-	Not detected	~46
259	c.364C>T	p.R122C	B	-	-	2.59	-	-	-	-	-	-	-	-	-	Not detected	~48
265	c.364C>T[§]	p.R122C[§]	B	-	-	2.59	-	-	-	-	-	-	-	-	-	Not detected	~59
293	c.613A>G	p.T205A	B	-	-	-	-	-	-	-	3.17	-	1	-	-	Not detected	~41
T473	c.986T>C	p.I329T	F	3.32	5.63	-	-	-	3.82	5.17	-	-	-	- (1ESS)	- (-)	Not detected	~50
14	c.1130G>A	p.C377Y	F	-	-	-	-	-	-	4.58	3.98	-	-	-	-	Not detected	~45
212	c.1177A>G	p.M393V	B	-	-	3.15	3.6	3.11	2.95	-	-	-	-	-	1	Not detected	46 (±1)
319	c.1850G>A	p.C617Y	B	-	-	-	-	-	-	-	2.80	-	2	-	-	Not detected	29 (±5)
319B	c.1850G>A[†]	p.C617Y[†]	B	-	-	-	-	-	-	-	2.80	-	2	-	-	Not detected	52 (±5)
254	c.1876G>A	p.G626R	B	-	-	-	-	3.69	3.12	-	3.02	-	-	-	-	Not detected	~53
454	c.1883G>T	p.C628F	F	-	-	-	-	-	-	6.14	3.83	-	-	-	-	Not detected	~50
412	c.1960G>A	p.D654N	A	-	-	-	-	-	-	-	-	-	-	-	-	**Skipping of exon 15**	-
262	c.1981_1982insT	p.C661L*fs*X20	B	-	-	-	-	-	-	-	-	-	-	- (1ESS)	- (-)	**Skipping of exon 16**	-
T472	c.2777delG	p.C926L*fs*X16	F	-	-	4.78	2.52	-	-	2.82	-	-	-	- (1ESS)	1 (-)	Not detected	ND
SV	c.2804G>A	p.C935Y	F	-	-	2.45	-	-	-	2.70	-	-	1	-	-	Not detected	~49
SS	c.2804G>A[†]	p.C935Y[†]	F	-	-	2.45	-	-	-	2.70	-	-	1	-	-	Not detected	~49
218	c.2956G>A	p.A986T	F	-	-	-	-	-	-	-	-	-	-	-	-	Not detected	~51
270	c.3116G>A	p.C1039Y	B	-	-	3.94	3.49	3.83	3.26	2.92	3.88	-	-	-	-	Not detected	~50
321	c.3149G>C	p.S1050T	B	-	-	-	-	-	-	-	-	-	-	-	-	Not detected	~55
WR1	c.3220T>C	p.C1074R	F	-	1.97	3.31	3.54	-	-	3.02	-	-	-	-	1	Not detected	~42
289	c.3250G>C	p.G1084R	B	3.53; -	4.14; 2.27	5.06	3.04	-	-	-	-	1	-	-	-	Not detected	~42
339	c.3503A>G	p.N1168S	F	-	-	2.93	3.38	4.15	3.99	-	-	-	-	-	-	Not detected	~52
252	c.3545G>A	p.C1182Y	B	-	-	-	-	-	4.25	2.92; -	-; 4.47	1	-	1	-	Not detected	~54

11	c.3751G>T	p.G1251C	F	-	-	-	-	-	-	-	-	-	- (1 ESS)	-	Not detected	~55
71	c.3761G>A	p.C1254Y	F	-	-	4.5; -	4.04; 2.85	3.51	2.94	-	-	-	-	-	Not detected	~56
380	c.3832T>A	p.C1278S	B	-	-	4.19	4.47	-	3.64	4.22	-	-	-	-	Not detected	~58
112	c.3886T>C	p.C1296R	F	-	3.66	4.42	5.18	-	-	-	-	-	-	2	-	Not detected	~48
381	c.3929G>T	p.G1310V	B	-	-	3.49	2.84	-	-	-	-	-	1	-	-	Not detected	~54
187	c.3974A>C	p.E1325A	F	-	-	-	-	-	3.67	-	3.76	1	-	- (1 ESS)	-	Not detected	~46
107	c.4025C>T	p.T1342I	F	2.10	-	-	-	2.74	-	4.10	-	-	-	-	-	Not detected	49 (±1)
89	c.4091_4095delinsCATTG	p.L1364_D1365deli	F	2.95	-	-	-	-	3.34	-	-	1	3	2	-	Not detected	~50
19	c.4121G>A	p.C1374Y	F	-	-	-	-	-	-	-; 4.71	3.61; 4.10	-	-	-	-	Not detected	~36
263	c.4143_4145delGAA	p.K1381del	B	-	-	3.28	-	-	-	-	-	3	-	1	-	Not detected	~23
114	c.4166G>A	p.C1389Y	B	-	-	3.14	2.69	-	-	-	-	-	-	-	1	Not detected	~54
117	c.4192G>A	p.D1398N	F	-	-	-	-	-	-	-	-	-	-	-	-	**5`cryptic splice site**	-
38	c.4217A>G	p.D1406G	B	-	-	-	-	-	-	-	-	2	-	-	-	Not detected	~63
382	c.4222T>C	p.C1408R	B	-	-	2.74	-	2.69	-	-	-	-	-	-	-	Not detected	~50
261	c.4270C>G	p.P1424A	B	2.35; -	2.10; 3.03	-	-	-	-	2.92	4.71	-	-	1	1	Not detected	~15
163	c.4306G>T	p.V1436L	F	-	-	-; 2.47	3.20; -	-	-	-	-	-	-	-	-	Not detected	49 (±2)
7	c.4337_4339delATA	p.D1446_I1447delin	F	-	-	-	-	-	-	-	-	-	-	-	-	Not detected	~43
KB	c.4409G>A	p.C1470Y	F	-	-	2.86	2.41	-	-	-	-	-	-	-	- (1ESS)	Not detected	~44
4B	c.4454G>A	p.C1485Y	F, B	-	-	-	-	-	2.97	-	-	1	2	-	-	Not detected	51 (±2)
39	c.4511A>G	p.N1504S	F	-	-	-	-	3.05; 3.88	3.61; -	-	-	-	-	-	-	Not detected	~44
78	c.4537T>C	p.C1513R	F	-	-	-	-	4.41	2.92	-	-	-	-	-	-	Not detected	~50
95	c.4588C>T	p.R1530C	F	-	-	-	-	2.92	-	-	-	-	-	-	3	Not detected	49 (±1)
395	c.5066A>G	p.D1689G	B	-	-	-	-	-	-	-	-	-	-	-	-	Not detected	~54
321	c.5084G>A	p.C1695Y	B	-	-	2.68	-	-	-	-	-	-	1	-	- (1 ESS)	Not detected	~55
121	c.5554G>A	p.E1852K	F	-	-	-	-	-	-	-	-	1	-	- (2 ESS)	-	Not detected	54 (±1)
73	c.5726T>C	p.I1909T	F	-	-	-	-	-; -	3.86; 3.46	-	-	-	2	-	-	Not detected	~46
260	c.5776_5778delAAT	p.N1926del	B	-	-	-	-	2.88	-	-	-	1	2	-	1	Not detected	NA
292	c.6069_6080del	p.C2024_G2027del	B	-	-	3.50; 2.48	-; -	-	3.99	-	3.16	-	-	1	-	Not detected	~49
459	c.6332G>T	p.C2111F	B	-	-	-	-	-	-	3.98	-	-	-	1	1	Not detected	~46
26	c.6577G>A	p.E2193K	F	2.44	-	-	-	-	-	-	-	1	1	1	3	Not detected	~4
T2966	c.6577G>A[§]	p.E2193K[§]	B	2.44	-	-	-	-	-	-	-	1	1	1	3	Not detected	~33
493	c.6610T>C	p.C2204R	F	2.13	4.66	-	2.47	3.48	5.19	-	2.93	-	-	-	-	Not detected	~44
150	c.6628T>C	p.C2210R	F, B	-	-	-	-	-	-	-	-	-	-	- (1 ESS)	-	Not detected	~45
9	c.7828G>A	p.E2610K	F	-	-	-	-	-	-	-	-	4	2	-	-	Not detected	~49
17	c.7852G>A	p.G2618R	F	5.29	3.22	-	-	-	-	4.96	3.17	2	2	2	2	Not detected	~53
173	c.8606T>C	p.L2869S	B	-	-	-	-	-	-	3.83	3.00	-	-	- (1 ESS)	-	Not detected	~53

Putative exonic polymorphisms																
T3078	c.510C>T	p.Y170Y	B	2.43	-	-	-	-	2.79	4.35	2.79	-	2	-	-	Not detected	~45
270	c.639C>T	p.V213V	B	3.17; -	-; 3.51	-	-	2.99	-	-	-	-	-	-	1	Not detected	~45
295B	c.1668G>A	p.V556V	B	-	-	-	-	-	-	4.37; 4.22	5.77; 3.61	-	-	-	- (1 ESS)	Not detected	~50
T3108	c.1845C>T	p.N615N	B	-	-	-	-	-	-	-	-	-	-	-	- (1 ESS)	Not detected	~70
345	c.2895G>A	p.E965E	B	-	-	-	-	-	-	-	-	1	4	4	2	Not detected	~53
373	c.3294C>T	p.D1098D	F	3.64	-	-	-	-	-	-	-	2	4	-	-	Not detected	~56
33	c.4428C>T	p.Y1476Y	B	-	-	2.98	-	-	-	-	-	-	-	-	- (1 ESS)	Not detected	~79
58	c.6354C>T	p.I2118I	F	-	-	2.68	3.32	-	-	-	-	-	-	-	-	**Skipping of exon 51**	-
245	c.6354C>T[§]	p.I2118I[§]	B	-	-	2.68	3.32	-	-	-	-	-	-	-	-	**Skipping of exon 51**	-
T473	c.6393C>T	p.C2131C	F	-	-	2.48	-	-	-	-	-	2	-	1 (1 ESS)	- (2 ESS)	Not detected	~47
150	c.6681A>C	p.S2227S	F, B	-	-	-; -	3.46; 3.77	-	-	-	-	-	-	-	-	Not detected	~49
150B	c.6681A>C[†]	p.S2227S[†]	B	-	-	-; -	3.46; 3.77	-	-	-	-	-	-	-	-	Not detected	~70
319	c.7209C>A	p.I2403I	B	-	-	-	-	-	-	-	-	-	1	-	- (1 ESS)	Not detected	71 (±5)
30263	c.8322G>A	p.K2774K	B	-	-	3.26	-	3.49	-	-	-	1	1	-	-	Not detected	~54

[a] Coded designation of investigated patients (two-letter code or numbers, also see Table 1).

[b] Sequence variants are described in relation to the translation initiation site of the mRNA reference sequence NM_000138.3. Mutation numbering is after HGVS and journal standards, with +1 as the A of the ATG initiation codon. Recurrent mutations in related (†) and unrelated (§) patients are marked.

[c] Material for transcript analysis: B, RNA extracted from PAXgene-stabilized whole blood; F, RNA extracted from primary dermal fibroblasts; A, RNA extracted from aortic tissue.

[d] Values represent high-scores for four different SR-proteins (values above the threshold). Threshold values provided by ESEfinder: SF2/ASF 1.956; SC35 2.383; SRp40 2.670; SRp55 2.676; -, no prediction (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi) [Cartegni et al., 2003].

[e] Number of putative exonic splicing enhancers predicted by RESCUE-ESE; -, no prediction (http://genes.mit.edu/burgelab/rescue-ese/) [Fairbrother et al., 2004].

[f] Number of putative exonic splicing enhancers predicted by PESX; -, no prediction. ESS in bracket denotes exonic splicing silencer predicted by PESX (http://cubweb.biology.columbia.edu/pesx/) [Zhang and Chasin, 2004; Zhang et al., 2005].

[g] Relative expression of the mutant allele revealed by quantitative sequence analyses according to the procedure outlined by [Magyar et al., 2009]. ND represents not detectable mutant allele. NA implies not applicable.

TABLE 3. *In vitro* and *in silico* Analyses of *FBN1* Intronic Sequence Variants Identified in This Study

#[a]	Sequence variant[c]	RNA source[b]	*In silico* analysis					*In vitro* cDNA analysis	
	Nucleotide change		HSF[d]	MaxEntScan[e]	NNSplice[f]	NetGene2[g]	ESEfinder 3.0[h]		
			dCV	dCV	Score change Wt-Mut	Score change Wt-Mut	Score change Wt-Mut	Aberrant splicing[i]	Effect on protein level
Canonical donor site substitutions									
498	c.164+1G>T	NA	-34%	-163%	-100%	-	-100%	-	-
24	c.1714+1G>T	F	-32%	-100%	-100%	-100%	-100%	**Exon 13 skipped (~2/4)**	**p.I531_D572del**
								49 nt of exon 13 deleted (~1/4)	**p.V556I*fs*X7**
								19 nt of intron 13 inserted (~1/4)	**p.D572G*fs*X10**
54	c.2728+1G>T	F	-34%	-111%	-100%	-100%	-	**Exon 22 skipped**	**p.P894_D910del**
84	c.3463+1G>C	F	-29%	-79%	-100%	-100%	-100%	**64 nt of exon 27 deleted (~1/2)**	**p.G1134T*fs*X7**
								63 nt of exon 27 deleted (~1/2)	**p.G1134_I1154del**
146	c.3838+1G>A	F	-28%	-92%	-100%	-100%	-100%	**Exon 30 skipping**	**p.I1239_D1280del**
277	c.5296+1G>A	B	-29%	-88%	-100%	-100%	-100%	**Exon 42 skipping**	**p.E1743_D1766del**
513	c.6037+2T>C	B	-30%	-82%	-100%	-100%	-100%	**78 nt of exon 48 deleted (~2/3)**	**p.G1987_E2012del**
								126 nt of intron 48 inserted (~1/3)	**p.D2013G*fs*X10**
94	c.6616+1G>C	F	-34%	-105%	-100%	-100%	-	**Exon 53 skipping (~1/2)**	**p.T2167_D2206del**
								34 nt of intron 53 inserted (~1/2)	**p.D2206G*fs*X15**
394	c.6997+1G>T	B	-29%	-92%	-100%	-100%	-100%	**Exon 56 skipping**	**p.E2292_D2333del**
T479	c.7453+1G>C	F	-29%	-89%	-100%	-100%	-100%	**Exon 59 skipping**	**p.N2446_L2486del**
213	c.7453+1G>A	NA	-32%	-90%	-100%	-100%	-100%	-	-
Canonical acceptor site substitutions									
28	c.1589-1_1591delins11	B	-100%	-100%	-100%	-100%	-	**11 nt of exon 13 deleted (~1/2)**	**p.D530V*fs*X10**
								20 nt of exon 13 deleted (~1/2)	**p.D530E*fs*X7**
30365	c.1715-2A>G	NA	-32%	-83%	-100%	-100%	-100%	-	-
3	c.1838-2A>G	F	-33%	-89%	-100%	-100%	-100%	**18 nt of intron 14 inserted**	**p.K612_D613ins6**
69B	c.2168-1G>C	F	-34%	-89%	-100%	-100%	-100%	**21 nt of exon 18 deleted**	**p.I724_D730del**
T31083	c.3590-1G>A	B	-36%	-104%	-100%	-100%	-100%	**Exon 29 skipping (~2/3)**	**p.S1202_C1242del**
								16 nt of exon 29 deleted (~1/3)	**p.D1197A*fs*X2**
154	c.4337-1G>A	NA	-5%;	-19%;	-100%	-100%	-100%	-	-
334	c.5789-1G>C	B	-32%	-101%	-100%	-100%	-100%	**Exon 47 skipped**	**p.V1931_D1973del**
410	c.6164-1G>A	F	-32%	-126%	-	-	-	**Exon 50 skipped**	**p.D2055_D2104del**
29203	c.6872-1G>T	B	-39%	-223%	-	-100%	-	**5 nt of exon 56 deleted**	**p.D2291E*fs*X2**
82	c.6998-1G>C	F	-32%	-111%	-100%	-100%	-100%	**12 nt of exon 57 deleted**	**p.D2333_E2336del**
231	c.7820-2A>T	B	-31%	-81%	-100%	-100%	-100%	**Exon 63 skipped**	**p.D2607G*fs*X68**
Other intronic variants at less or non-conserved positions									

31326	c.165-86C>A	F	-3%	+213%, -1%	-100%	-	-	Not detected	-
31326	c.248-44G>A	F	-	-	-	-	-	Not detected	-
284	c.346+26A>G	NA	-	-	-	-	-	-	-
506	c.347-73A>G	B	-0.1%	-38%	-	-	-	Not detected	-
305	c.539-138C>T	NA	-9%; +2%	-	-	-	-	-	-
284	c.737-26delT	NA	-	-925%	-	-	-	-	-
T30783	c.737-32T>G	B	-	-21%	-	-	-	Not detected	-
459	c.863-10dupT	B	-62%; +352%	-163%; +186%	-2%	-	-	Not detected	-
29276	c.863-121T>G	B	-	-	-	-	-	Not detected	-
287	c.1468+5G>C	B	-14%	-80%	-100%	-100%	-	**18 nt of intron 11 inserted (~1/3)** **48 nt of exon 11 deleted (~1/3)** **50 nt of exon 11 deleted (~1/3)**	**p.D490G*fs*X3** **p.C474_D490delinsY** **p.E475X**
75	c.1714+37A>G	F	-	-	-	-	-	Not detected	-
T30783	c.1837+16C>T	B	-	-	-	-	-	Not detected	-
352	c.2678-70C>G	B	-13%	+11%; -108%	-	-	-	Not detected	-
258	c.3964+61A>G	B	-	-	-	-	-	Not detected	-
175	c.4337-29T>A	NA	2%	-	-	-	-	-	-
31174	c.4583-5A>G	B	-0.09%	-257%	+100%	+100%	-	**4 nt of intron 36 inserted**	**p.D1528A*fs*X52**
30964	c.4747+82T>C	B	-	-	-	-	-	Not detected	-
65	c.5225-13T>G	F	-	+1%	-	-	-	Not detected	-
249	c.5225-19C>T	NA	-	-0.2%	-	-	-	-	-
289	c.5296+30C>A	B	-1%	-	-	-	-	Not detected	-
6	c.5671+3_5671+4insCC	F	-30%	-223%	-100%	-100%	-100%	**Exon 45 skipped**	**p.R1850_D1891del**
358	c.5788+10C>A	NA	-0.2%; +15%	-60	-	+1%	-	-	-
459	c.6037+40_50del11	B	-81%	-	-	-	-	Not detected	-
514	c.6313+53A>G	NA	+66%	-	-	+15%; +11%	-2%; +2%	-	-
52	c.6314-15G>A	B	-41%	+13%	-	-	-4%	**Exon 51 skipped**	**p.E2105_V2126del**
463	c.6496+5G>C	NA	-12%; -40%	-58%; +453%	-100%	-	-100%	-	-
188	c.6997+4_6997+13delinsT	F	-22%	-79%	-100%	-100%	-100%	**51 nt of exon 56 deleted**	**p.C2316_D2333delinsY**
383	c.7204+7C>G	B	-	-	-3%	-6%	-	Not detected	-
281B	c.7204+54dupT	NA	-98%; +497%;	+105%; -406%;	-	-	-	-	-
152	c.7330+3_7330+8delins15	F	-29%	-100%	-100%	-100%	-100%	**Exon 58 skipped**	**p.I2403_D2444del**
88	c.7699+5delG	F, B	-16%	-97%	-100%	-100%	-100%	**30 nt of intron 61 inserted (~2/3)** **140 nt of intron 61 inserted (~1/3)**	**p.E2566_D2567ins10** **p.D2567G*fs*X32**

88B	c.7699+5delG†	F, B	-16%	-97%	-100%	-100%	-100%	**30 nt of intron 61 inserted (~2/3)**	**p.E2566_D2567ins10**
								140 nt of intron 61 inserted (~1/3)	**p.D2567G*fs*X32**
30205	c.7699+5G>A	B	-14%	-61%	-100%	-100%	-32%	**31 nt of intron 61 inserted (~1/2)**	**p.D2567G*fs*X11**
								141 nt of intron 61 inserted (~1/2)	**p.D2567G*fs*X11**
371	c.8052-23T>C	NA	-	-	-	-	-	-	-

[a] Coded designation of investigated patients (two-letter code or numbers, also see Table 1).

[b] Material for transcript analysis: B, RNA extracted from PAXgene-stabilized whole blood; F, RNA extracted from primary dermal fibroblasts; NA, material not available for transcript analysis.

[c] Sequence variants are described in relation to the translation initiation site of the mRNA reference sequence NM_000138.3. Mutation numbering is after HGVS and journal standards, with +1 as the A of the ATG initiation codon. Recurrent mutation in related (†) patient is marked.

[d] Delta consensus value (dCV) calculates the difference between the wild-type and mutant consensus value predicting by Human Splicing Finder version 2.4 (HSF); (http://www.umd.be/HSF/) [Desmet et al., 2009].

[e] Delta consensus value (dCV) calculates the difference between the wild-type and mutant consensus value predicting by the MaxEntScan algorithm, which is the part of Human Splicing Finder version 2.4 interface (HSF); (http://www.umd.be/HSF/) [Yeo and Burge, 2004; Desmet et al., 2009].

[f] Splice site scores predicted by NNSplice, the output of the network is a score between 0 and 1 for a potential splice site. Scores close to 1 indicate strong splice site, while values close to 0.4 (threshold) indicate weak prediction (http://www.fruitfly.org/seq_tools/splice.html) [Reese et al., 1997].

[g] Splice site scores predicted by NetGene2, the output of the network is a score between 0 and 1 for a potential splice site. Scores close to 1 indicate strong splice site, while values close to 0 indicate weak prediction (http://www.cbs.dtu.dk/services/NetGene2/) [Brunak et al., 1991].

[h] Splice site scores predicted by ESEfinder, scores above the threshold mean potential splice site. Threshold values provided by ESEfinder: 6.67 for 5` splice site and 6.632 for 3` splice site (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi) [Cartegni et al., 2003].

[i] Values in parentheses indicate the approximately ratio of splice products by comparing to each other.

TABLE 4. Summary of *in vitro* and *in silico* Analyses of *FBN1* Exonic Variants

in vitro / *in silico*	ESEfinder 3.0		RESCUE-ESE		PESX	
	Predicted[a]	**Not predicted[b]**	**Predicted[a]**	**Not predicted[b]**	**Predicted[a]**	**Not predicted[b]**
Aberrant splicing	1	3	0	4	0	4
Normal splicing	47	19	18	48	10	56

[a] Variants, which were predicted to lie within a putative ESE motif.

[b] Variants, which were not predicted to lie within a putative ESE motif.

(for detailed information about the used programs see the description of Table 2).

TABLE 5. Sensitivity, Accuracy, and Specificity of Three Exonic Splicing Enhancer Prediction Algorithms

	Exonic Variants[a]		
	ESEfinder 3.0	**RESCUE-ESE**	**PESX**
Sensitivity[b] (%)	25	NA	NA
Accuracy[c] (%)	29	69	80
Specificity[d] (%)	29	73	93

[a] Exonic variants analyzed *in silico* and *in vitro* in this study (see Table 2).

[b] Sensitivity was calculated by the following formula: number of true positives/number of true positives+number of false negatives.

[c] Accuracy was calculated by the following formula: number of true positives+number of true negatives/number of true positives+number of false positives+number of true negatives+number of false negatives.

[d] Specificity was calculated by the following formula: number of true negatives/number of true negatives+number of false positives.

NA implies not applicable.

TABLE 6. Summary of *in vitro* and *in silico* Analyses of *FBN1* Intronic Variants

A. Canonical Donor Site Substitutions

in vitro / *in silico*	Human Splicing Finder		MaxEntScan		NNSplice		NetGene2		ESEfinder 3.0	
	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]
Aberrant splicing	9	0	9	0	9	0	9	0	7	2
Normal splicing	0	0	0	0	0	0	0	0	0	0

B. Canonical Acceptor Site Substitutions

in vitro / *in silico*	Human Splicing Finder		MaxEntScan		NNSplice		NetGene2		ESEfinder 3.0	
	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]
Aberrant splicing	9	0	9	0	7	2	8	1	6	3
Normal splicing	0	0	0	0	0	0	0	0	0	0

C. Other Intronic Variants at Less- or Non-conserved Positions

in vitro / *in silico*	Human Splicing Finder		MaxEntScan		NNSplice		NetGene2		ESEfinder 3.0	
	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]	Predicted[a]	Not predicted[b]
Aberrant splicing	7	1	7	1	7	1	7	1	5	3
Normal splicing	5	10	4	11	1	14	0	15	0	15

[a] Variants, which were predicted to lie within a putative splice site motif.

[b] Variants, which were not predicted to lie within a putative splice site motif.

(for detailed information about the used programs see the description of Table 3).

TABLE 7. Sensitivity, Accuracy, and Specificity of Five Splice Site Prediction Algorithms

A. Canonical Donor Site Substitutions

	Canonical Donor Site[a]				
	Human Splicing Finder	**MaxEntScan**	**NNSplice**	**NetGene2**	**ESEfinder 3.0**
Sensitivity[d] (%)	100	100	100	100	77
Accuracy[e] (%)	100	100	100	100	77
Specificity[f] (%)	NA	NA	NA	NA	NA

B. Canonical Acceptor Site Substitutions

	Canonical Acceptor Site[b]				
	Human Splicing Finder	**MaxEntScan**	**NNSplice**	**NetGene2**	**ESEfinder 3.0**
Sensitivity[d] (%)	100	100	77	89	67
Accuracy[e] (%)	100	100	77	89	67
Specificity[f] (%)	NA	NA	NA	NA	NA

C. Other Intronic Variants at Less- or Non-conserved Positions

	Other Intronic Position[c]				
	Human Splicing Finder	**MaxEntScan**	**NNSplice**	**NetGene2**	**ESEfinder 3.0**
Sensitivity[d] (%)	88	88	88	86	63
Accuracy[e] (%)	74	78	91	96	87
Specificity[f] (%)	67	73	93	100	100

[a] Canonical donor site denotes the +1 and +2 positons of the 5` splice site analyzed *in silico* and *in vitro* in this study (see Table 3).

[b] Canonical acceptor site denotes the -1 and -2 positons of the 3` splice site analyzed *in silico* and *in vitro* in this study (see Table 3).

[c] Other intronic position denotes positions different from the canonical donor and acceptor sites analyzed *in silico* and *in vitro* in this study (see Table 3).

[d] Sensitivity was calculated by the following formula: number of true positives/number of true positives+number of false negatives.

[e] Accuracy was calculated by the following formula: number of true positives+number of true negatives/number of true positives+number of false positives+number of true negatives+number of false negatives.

[f] Specificity was calculated by the following formula: number of true negatives/number of true negatives+number of false positives.

NA implies not applicable.

TABLE 6. RSCU Values for 9 Silent Mutations Identified in This Study

#[a]	SNP[b]	Wt-codon	Mut-codon	Wt RSCU value[c]	Mut RSCU value[c]	dRSCU values (%)[d]
270	c.639C>T	GTC	GTT	14.5	11.0	-6
295B	c.1668G>A	GTG	GTA	28.1	7.1	-34
345	c.2895G>A	GAG	GAA	39.6	29.0	-9
373	c.3294C>T	GAC	GAT	25.1	21.8	-3
33	c.4428C>T	TAC	TAT	15.3	12.2	-10
58	c.6354C>T	ATC	ATT	20.8	16.0	-7
150	c.6681A>C	TCA	TCC	12.2	17.7	+9
319	c.7209C>A	ATC	ATA	20.8	7.5	-20
408	c.8322G>A	AAG	AAA	31.9	24.4	-9

[a] Coded designations of investigated patients.

[b] SNPs are described in relation to the translation initiation site of the *FBN1* mRNA reference sequence NM_000138.3. Mutation numbering is after HGVS and journal standards, with +1 as the A of the ATG initiation codon.

[c] Relative Synonymous Codon Usage (RSCU) values represent the frequency of the given synonymous codon for *Homo sapiens* resulted from the Codon Usage Database (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=9606)

[d] dRSCU represents the difference between mutant and wild-type RSCU values in percentage.

FIGURE 1.

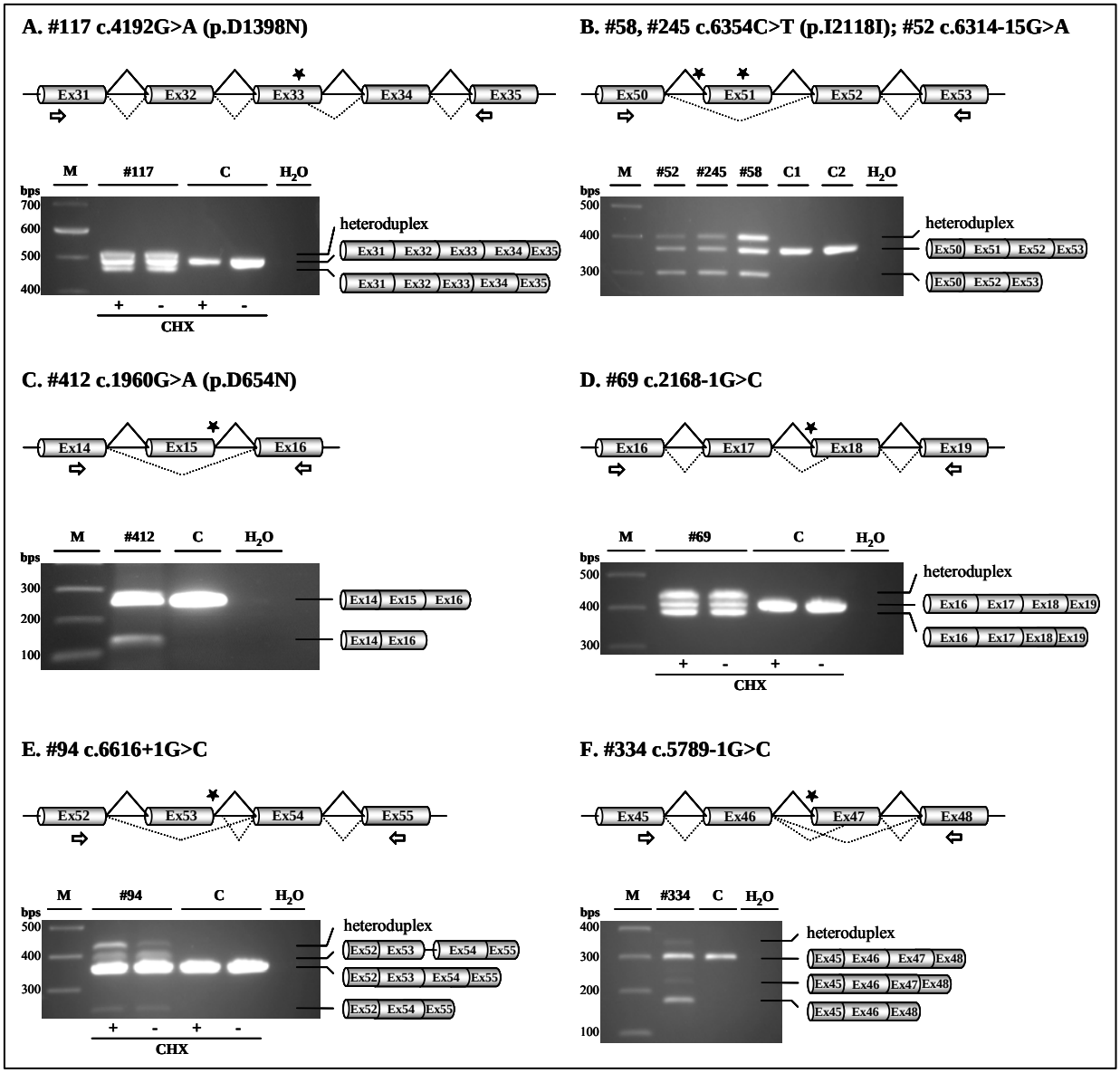

FIGURE 1. RT-PCR analyses of *FBN1* exonic and intronic variants affecting splicing. Schematic diagrams display the exon-intron arrangement of the *FBN1* genomic segments tested for cDNA analysis. Boxes with number represent exons and introns are indicated as horizontal lines between boxes. The position of the mutation is marked by black asterisk. Fragments for cDNA analyses are amplified by OneStep RT-PCR using primers indicated as open arrows (⇨ for forward, and ⇦ for reverse directions). Normal pre-mRNA splicing is represented by black lines, while aberrant splicing by dashed lines. The splicing patterns are displayed schematically alongside the right part of the agarose gel pictures. If RNA from fibroblasts was used, prior to RNA extraction one subculture of fibroblast cell lines was treated (+) with the translation-inhibitor cycloheximide (CHX), while to one subculture no CHX (-) was added. M denotes the molecular weight marker. **A:** RT-PCR analysis of the exonic mutation c.4192G>A (p.D1398N) of Patient 117. The sequence change was localized at nucleotide position 105 from 123 of exon 33. The deletion of 21 bp at the end of exon 33 was observed (lower band) in the mutation-carrying sample (#117), which was absent in the control sample (C). The

uppermost band proved to be a heteroduplex of the wild-type and mutant transcript, as confirmed by sequencing. **B:** RT-PCR analysis of the exonic c.6354C>T (p.I2118I) silent mutation (Patients 58 and 245) and the intronic variant c.6314-15G>A (Patient 52). The sequence alteration affects the 41 nucleotide from 66 of exon 51. The skipping of exon 51 was detected in the case of both mutations. The exon-skipped transcript and the heteroduplex band are not present in the control samples (C1 is a control RNA isolated from blood and C2 is a control RNA extracted from fibroblasts. Note that the splicing defect was identical both in fibroblasts (#58) and PAXgene blood sample (#245). **C:** RT-PCR analysis of the c.1960G>A (p.D654N) mutation of Patient 24. This mutation caused the skipping of exon 15. The control sample displays the normal wild-type band. **D:** RT-PCR analysis of the c.2168-1G>C acceptor site mutation of Patient 69 revealed two additional fragments compared two the control probe, the deletion of the first 21 bp of exon 18 and a heteroduplex. **E:** RT-PCR analysis of the c.6616+1G>C donor site substitution in Patient 94 resulted in complex aberrant splicing. Two aberrant transcripts were observed, the skipping of exon 53 and the inclusion of the first 28 bp from intron 53 to exon 53. Additionally, a heteroduplex was detected. **F:** RT-PCR analysis of the c.5789-1G>C acceptor site mutation of Patient 334. Skipping of exon 47, exclusion of the first 77 bp of exon 47, and a heteroduplex were detected in the mutation-carrying samples. These fragments were absent in the control.

FIGURE 2.

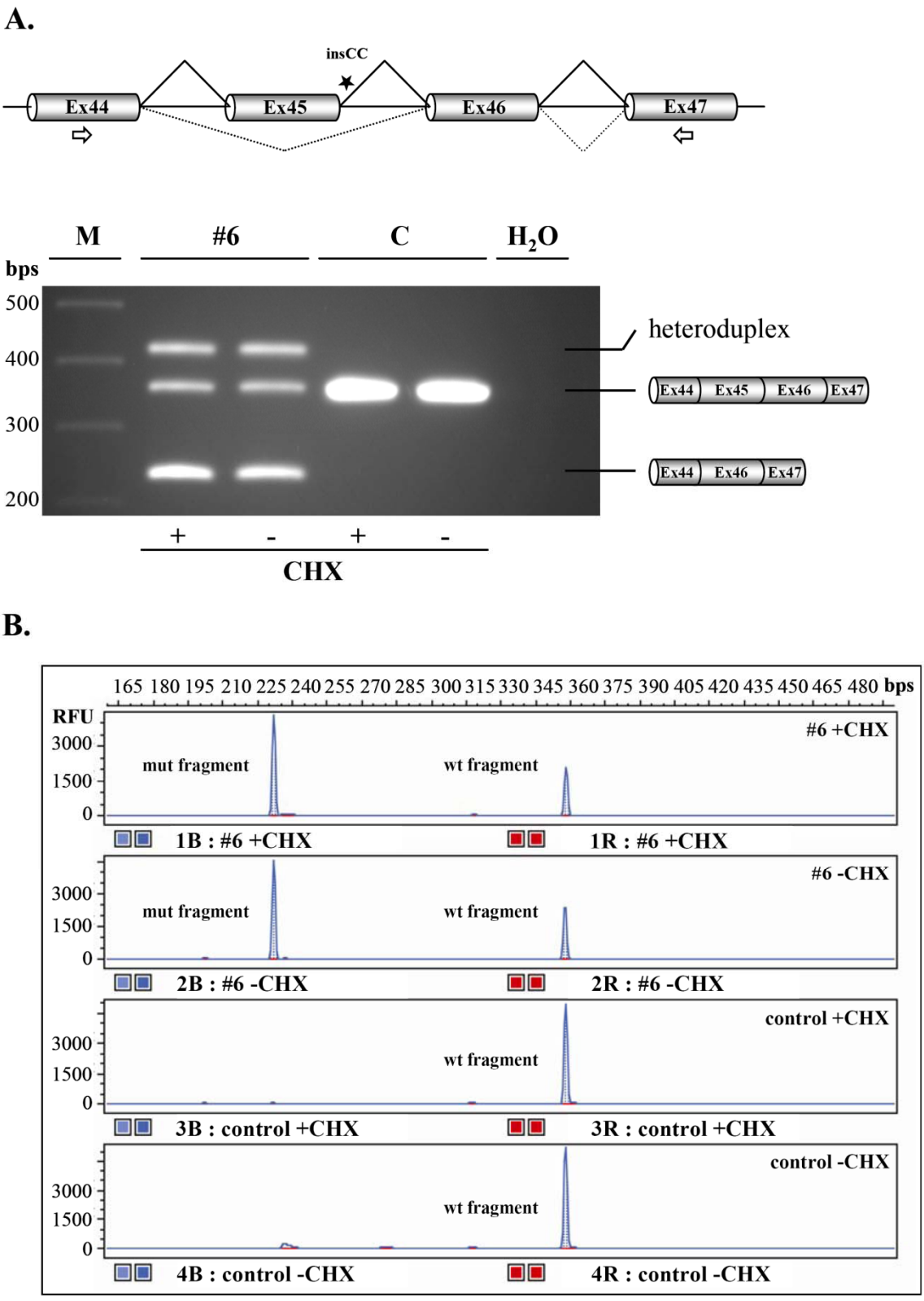

FIGURE 2. RT-PCR and fragment analyses of the intronic mutation c.5671+3_5671+4insCC (Patient 6). **A:** Agarose gel electrophoresis of RT-PCR products. Schematic diagram shows the exon-intron arrangement of the *FBN1* gene from exon 44-47. Figure legend and symbols are as described in Figure 1. This mutation revealed

the skipping of the entire exon 45. Additionally, together with the wild-type transcript formed a heteroduplex. Both bands are absent in the control sample. **B:** Fragment analysis with fluorescently labeled amplicons using ABI310 Genetic Analyzer. Note that the denaturing condition resulted in the exon-skipped fragment and the wild-type fragment, but not the band proved to be the heteroduplex.

FIGURE 3.

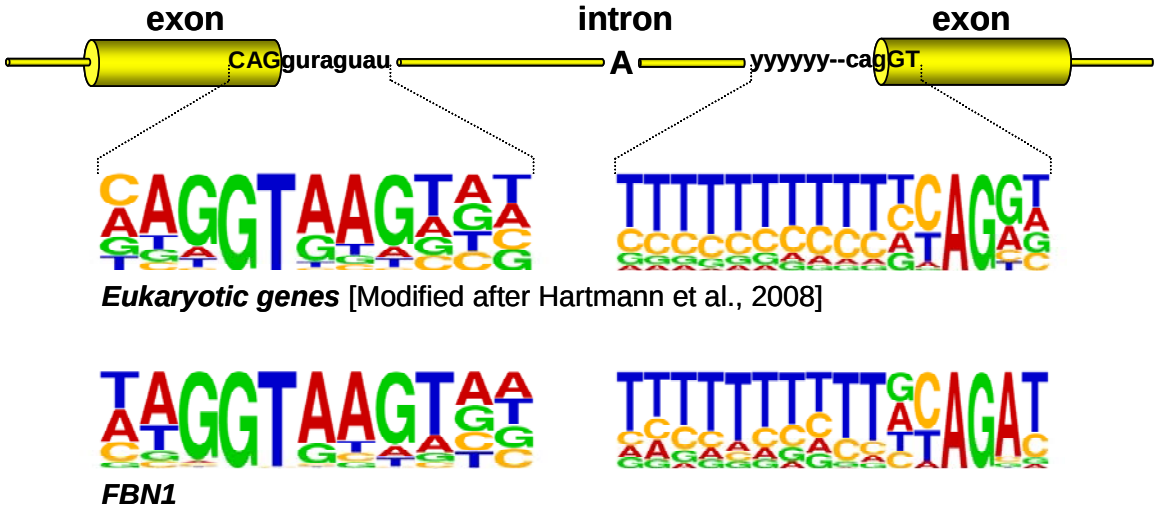

FIGURE 3. Schematic representation of exon-intron boundaries of a eukaryotic gene and the nucleotide usage of the *FBN1* gene at those positions (sequence logos were made by the sequence logo generator WebLogo3 http://weblogo.berkeley.edu/) [Schneider et al., 1990; Crooks et al., 2004].

Contributions of authors to the manuscript "*FBN1* pre-mRNA Splicing Alterations: Comparative *in vitro* Transcript Analyses and *in silico* Prediction for the Effect of *FBN1* Exonic and Intronic Mutations"

István Magyar	Planning and performing experiments, preparation of RNA, primer design, RT-PCR, sequencing, interpretation of data, writing of the manuscript
Eliane Arnold	Mutation analyses, technical assistance
Dvora Colman	Preparation of RNA, RT-PCR, sequencing
Janine Meienberg	Technical assistance in RT-PCR and sequencing analyses
Daniela Baumgartner	Sending of patient`s material
Armand Bottani	Sending of patient`s material
Siv Fokstuen	Sending of patient`s material
Marie-Claude Addor	Sending of patient`s material
Thierry Carrel	Sending of patient`s material
Beat Steinmann	Sending of patient`s material
Wolfgang Berger	Supervision of the study, conceptual input, interpretation of data, editing of the manuscript
Gábor Mátyás	Conceptual planning, supervision of the study

3.2 Quantitative Sequence Analysis of *FBN1* Premature Termination Stop Codons Provide Evidence for Incomplete NMD in Leukocytes

Manuscript published in Human Mutation, September 2009, 30(9):1355-1364

István Magyar,[1] Dvora Colman,[1] Eliane Arnold,[1,2] Daniela Baumgartner,[3] Armand Bottani,[4] Siv Fokstuen,[4] Marie-Claude Addor,[5] Wolfgang Berger,[1] Thierry Carrel,[6] Beat Steinmann,[2] and Gábor Mátyás[1]

[1] Division of Medical Molecular Genetics and Gene Diagnostics, Institute of Medical Genetics, University of Zurich, Zurich, Switzerland;

[2] Division of Metabolism and Molecular Pediatrics, University Children's Hospital, Zurich, Switzerland;

[3] Department of Pediatric Cardiology, Innsbruck Medical University, Innsbruck, Austria;

[4] Division of Medical Genetics, Geneva University Hospitals, Geneva, Switzerland;

[5] Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland;

[6] Clinic for Cardiovascular Surgery, University Hospital, Berne, Switzerland

Correspondence to:

Gábor Mátyás, Ph.D., Division of Medical Molecular Genetics and Gene Diagnostics, Institute of Medical Genetics Institute of Medical Molecular Genetics, University of Zurich, Schorenstrasse 16, CH-8603 Schwerzenbach, Switzerland; e-mail: matyas@medgen.uzh.ch

METHODS

Human Mutation

OFFICIAL JOURNAL

HGVS HUMAN GENOME VARIATION SOCIETY

www.hgvs.org

Quantitative Sequence Analysis of *FBN1* Premature Termination Codons Provides Evidence for Incomplete NMD in Leukocytes

István Magyar,[1] Dvora Colman,[1] Eliane Arnold,[1,2] Daniela Baumgartner,[3] Armand Bottani,[4] Siv Fokstuen,[4] Marie-Claude Addor,[5] Wolfgang Berger,[1] Thierry Carrel,[6] Beat Steinmann,[2] and Gábor Mátyás[1]*

[1]*Division of Medical Molecular Genetics and Gene Diagnostics, Institute of Medical Genetics, University of Zurich, Zurich, Switzerland;* [2]*Division of Metabolism and Molecular Pediatrics, University Children's Hospital, Zurich, Switzerland;* [3]*Department of Pediatric Cardiology, Innsbruck Medical University, Innsbruck, Austria;* [4]*Division of Medical Genetics, Geneva University Hospitals, Geneva, Switzerland;* [5]*Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland;* [6]*Clinic for Cardiovascular Surgery, University Hospital, Berne, Switzerland*

Communicated by Mats Nilsson

Received 18 December 2008; accepted revised manuscript 15 May 2009.

Published online 22 May 2009 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/humu.21058

ABSTRACT: We improved, evaluated, and used Sanger sequencing for quantification of single nucleotide polymorphism (SNP) variants in transcripts and gDNA samples. This improved assay resulted in highly reproducible relative allele frequencies (e.g., for a heterozygous gDNA 50.0±1.4%, and for a missense mutation-bearing transcript 46.9±3.7%) with a lower detection limit of 3–9%. It provided excellent accuracy and linear correlation between expected and observed relative allele frequencies. This sequencing assay, which can also be used for the quantification of copy number variations (CNVs), methylations, mosaicisms, and DNA pools, enabled us to analyze transcripts of the *FBN1* gene in fibroblasts and blood samples of patients with suspected Marfan syndrome not only qualitatively but also quantitatively. We report a total of 18 novel and 19 known *FBN1* sequence variants leading to a premature termination codon (PTC), 26 of which we analyzed by quantitative sequencing both at gDNA and cDNA levels. The relative amounts of PTC-containing *FBN1* transcripts in fresh and PAXgene-stabilized blood samples were significantly higher (33.0±3.9% to 80.0±7.2%) than those detected in affected fibroblasts with inhibition of nonsense-mediated mRNA decay (NMD) (11.0±2.1% to 25.0±1.8%), whereas in fibroblasts without NMD inhibition no mutant alleles could be detected. These results provide evidence for incomplete NMD in leukocytes and have particular importance for RNA-based analyses not only in *FBN1* but also in other genes.
Hum Mutat 30:1355–1364, 2009. © 2009 Wiley-Liss, Inc.

KEY WORDS: Marfan syndrome; fibrillin-1; nonsense mutation; nonsense-mediated mRNA decay; transcripts in blood; PAXgene blood system; quantitative sequencing

Introduction

Nonsense-mediated messenger RNA (mRNA) decay (NMD) is an evolutionary conserved quality control mechanism to detect a premature termination (nonsense) codon (PTC) and prevent the expression of truncated proteins [Maquat, 1995; Mühlemann et al., 2008]. In mammalian cells, NMD is triggered by a postsplicing exon-junction complex (EJC) of proteins that is downstream of a PTC and, hence, is not removed during translation. NMD consists of a series of steps that ultimately lead to the degradation of mRNA, resulting in functional haploinsufficiency [Chang et al., 2007; Stalder and Mühlemann, 2008; Wilkinson, 2005]. In contrast, transcripts containing a PTC near the initiation AUG codon or located downstream from the last ~55 bp of the penultimate exon of a gene escape NMD, resulting in the synthesis of truncated proteins with different lengths that can be pathogenic through dominant-negative or gain-of-function effects [e.g., Inacio et al., 2004; Sanchez-Sanchez et al., 2007]. Consequently, NMD cannot only prevent the accumulation of harmful proteins but can also modulate the clinical manifestations of genetic disorders, which has been reported by several previous studies [e.g., Dietz et al., 1993a; Frischmeyer and Dietz, 1999; Karttunen et al., 1998; Khajavi et al., 2006; Mort et al., 2008; Schrijver et al., 2002].

Due to the preferentially degraded and thus often undetectable amount of the mutant transcript, NMD can also hamper the detection of heterozygous PTC mutations at the RNA level. This is often the case in molecular genetic testing because roughly one-third of all disease-causing mutations in humans induce premature termination of translation [e.g., Beroud et al., 2000; Mendell and Dietz, 2001]. In addition, RNA-based mutation detection is often hindered by the lack of appropriate samples, in which the NMD pathway can be pharmacologically inhibited prior to RNA extraction, for example, by using translation-inhibiting drugs (anisomycin, cycloheximide, emetine, pactamycin, and puromycin) which increase the abundance of PTC-containing transcripts [Andreutti-Zaugg et al., 1997; Noensie and Dietz, 2001]. For these reasons, mutation screening is often performed at the genomic DNA (gDNA) level, rather than on RNA, even if RNA-based approaches allow efficient screening of the entire coding region of a large gene and permit identification of certain mutations undetectable by standard exon-by-exon gDNA screening.

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Gábor Mátyás, Division of Medical Molecular Genetics and Gene Diagnostics, Institute of Medical Genetics, University of Zurich, Schorenstrasse 16, CH-8603 Schwerzenbach, Switzerland. E-mail: matyas@medgen.uzh.ch

Nevertheless, there is an increasing need for the analysis of mRNA processing steps because such investigations are important to assess the pathogenic consequence of mutations identified at gDNA level. Indeed, there is a growing number of examples demonstrating that not only nonsense and frameshift, but also silent, missense, and in-frame mutations can result in PTC-containing transcripts and that novel intronic sequence changes may have unpredictable effects on splicing, thus increasing the demand for mutation-oriented RNA analyses in the large number of sequence variants with unknown effect [Caputi et al., 2002; Cartegni et al., 2002; Liu et al., 1997a]. Commercially available systems for RNA extraction from easily accessible blood samples, such as the PAXgene Blood RNA System (PreAnalytix; www.preanalytix.com), meet this increasing demand for transcript analyses. However, little is known about the efficiency of NMD in blood compared to tissues affected by the underlying disease.

In this study, we address this issue by analyzing transcripts of the fibrillin-1 (*FBN1*; MIM# 134797) gene from patients with suspected Marfan syndrome (MFS; MIM# 154700), an autosomal dominant connective tissue disorder, which displays variable manifestations in the cardiovascular, ocular, and skeletal systems and is caused by *FBN1* mutations in the majority of cases [De Paepe et al., 1996; Dietz et al., 1991]. We report 18 novel and 19 known *FBN1* sequence variants leading to PTC, 26 of which we studied by improved quantitative sequencing both at the genomic and RNA levels. Comparative analysis of transcripts derived from affected primary dermal fibroblasts with and without NMD inhibition as well as from fresh and PAXgene-stabilized blood samples demonstrated tissue-specific degradation of PTC-containing *FBN1* transcripts, providing evidence for incomplete NMD in white blood cells (leukocytes). The advantages and disadvantages of RNA analyses in leukocytes are discussed.

Materials and Methods

Patients and Detection of Mutations Leading to PTC

We have performed *FBN1* mutation analysis in nearly 400 unrelated patients with suspected MFS using our previously described mutation screening strategy [Mátyás et al., 2002a]. We have detected 135 *FBN1* sequence variants, comprising missense (55%), nonsense (12%), and splice site mutations (18%), as well as small deletions (10%), insertions (1%), and insertion-deletions (4%), among which 37 were expected to introduce a PTC (Table 1). Sequence variant nomenclature follows the guidelines of the Human Genome Variation Society (HGVS, www.hgvs.org/mutnomen), with numbering using +1 as A of the ATG start codon of the *FBN1* mRNA reference sequences NM_000138.3.

Data on the clinical phenotypes of patients were collected from medical records or during physical examinations by one of the authors (D.B., A.B., S.F., M.-C.A., T.C., and B.S.).

Cell Culture

Fibroblasts from skin biopsies were cultured in Minimal Essential Medium (MEM) with Earle's salts without L-glutamine (PAA Laboratories GmbH, Cölbe, Germany; www.paa.com) supplemented with 10% fetal bovine serum (PAA Laboratories GmbH, Cölbe, Germany; www.paa.com), 1 × antibiotic–antimycotic (Gibco, Invitrogen Corporation, Paisley, UK; www.invitrogen.com), and 2 mM L-glutamine (LabForce AG, Nunningen, Switzerland; www.labforce.ch) and incubated at 37°C with 5% CO_2. Prior to RNA extraction, each cell culture at ~100% confluence was split and one subculture (+CHX) was treated with the translation-inhibitor cycloheximide (Actidione; Sigma, St. Louis, MO; www.sigmaaldrich.com) for 6–8 hr at a concentration of 100 µg/ml to inhibit NMD, while to one subculture (−CHX) no cycloheximide was added.

DNA and RNA Extraction

Total DNA was extracted from fibroblasts and EDTA-anticoagulated whole blood samples by standard procedures or was referred to us for mutation detection in *FBN1*.

Total RNA was isolated from both cycloheximide-treated and untreated fibroblast subcultures by using the RNeasy mini kit (Qiagen, Hilden, Germany; www.qiagen.com) and/or from PAXgene-stabilized whole blood using the PAXgene blood RNA kit (Qiagen, Hilden, Germany; www.qiagen.com). For Patient BP, total RNA was also isolated by the RNeasy mini kit (Qiagen, Hilden, Germany; www.qiagen.com) from leukocytes derived immediately from EDTA-anticoagulated whole blood (1 volume) after sample collection and lysis of erythrocytes for 10 min at 4°C by 8 volumes isoosmotic ammonium chloride solution (0.839% ammonium chloride, 0.1% potassium bicarbonate, and 0.004% EDTA Na_2). RNAs of Patients 101 and 344 were treated with two units TURBO DNase (Ambion, Applied Biosystems, Rotkreuz, Switzerland; www.appliedbiosystems.com) to assess the effect of this treatment on quantitative sequencing.

DNA and RNA concentrations were determined by using a NanoDrop ND-1000 system (NanoDrop Technologies, Wilmington, DE; www.nanodrop.com) according to the manufacturer's instructions. RNA qualities (RNA integrity number, RIN) were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA; www.agilent.com). RNA samples with RIN >4 were used as a template for transcript analyses.

Qualitative Transcript Analyses

Prior to quantitative sequencing, qualitative transcript analysis was performed to investigate the effect of mutations on pre-mRNA splicing. Reverse transcription (RT)-PCR was used to amplify complementary DNA (cDNA) fragments containing a PTC-introducing mutation. RT-PCR was performed by means of the Qiagen OneStep RT-PCR kit (Qiagen, Hilden, Germany; www.qiagen.com) using 125 ng total RNA and cDNA-specific primers flanking the respective mutation (Supp. Table S1) with the following cycling profile: 50°C for 30 min, 95°C for 15 min, followed by 32–35 (fibroblast RNA) or 40 (blood RNA) cycles of 94°C for 30 sec, annealing temperature (Supp. Table S1) for 45 sec, and 72°C for 30 sec, followed by a 10-min final extension step at 72°C. RT-PCR products (0.8–2.0 µl) were purified using 0.1 µl ExoSAP-IT (USB Corporation, Cleveland, OH; www.usbweb.com) in a total volume of 6 µl by incubating at 37°C for 30 min and at 80°C for 15 min. Purified RT-PCR products were sequenced directly in both directions on an ABI PRISM 3100 or 3730 Genetic Analyzer with POP-6 or POP-7 polymer (ABI3100, ABI3730, POP6, POP7; Applied Biosystems, Rotkreuz, Switzerland; www.appliedbiosystems.com) using 0.8 µM of primer, 1.5 µl of 5 × sequencing buffer, and 1 µl of BigDye terminator v1.1 (Applied Biosystems, Rotkreuz, Switzerland; www.appliedbiosystems.com) in a final reaction volume of 10 µl and subsequently purified using Sephadex G-50 according to the manufacturer's recommendations (Millipore Technical Note TN053; www.millipore.com). To detect transcripts of abnormal length, RT-PCR products were also separated and visualized by standard gel electrophoresis using ethidium bromide-containing 1.5% agarose gels for 1–2 hr at 7.7 V/cm. All thermal cycles were run

Table 1. PTC-Introducing *FBN1* Mutations Identified in This Study

Patient								Sequence variant[d]			
			Phenotype[c]								
#[a]	Year of birth	Material[b]	SS	OS	CS	FH	Disorder	Location	Nucleotide change	AA change	Reference
211	1973	B_P	(+)	–	+	+	MFS	Exon 2	c.229G>T	p.G77X	This study
59	1960	B_P, F	+	?	(+)	–	MFS*	Intron 2	c.247+1G>A (c.165_247del)	p.P56H*fs*X45	[Dietz et al., 1993a]
46	1966	B_P, F	+	–	+	+	MFS	Intron 3	c.347−2A>G (c.346_347insG)	p.I116S*fs*X13	This study
46B	1993	B_P	(+)	–	?	+	MFS	Intron 3	c.347−2A>G (c.346_347insG)†	p.I116S*fs*X13†	This study
18	1990	NA	+	–	+	+	MFS	Exon 5	c.508delT	p.Y170T*fs*X20	[Baumgartner et al., 2005]
21	1977	NA	+	–	+	–	MFS*	Exon 6	c.651G>A	p.W217X	[Baumgartner et al., 2005]
257	1982	NA	?	?	?	?	MFS**	Exon 8	c.917delA	p.N306T*fs*X24	This study
16	1997	F	+	–	+	–	MFS	Exon 10	c.1206delT	p.P404H*fs*X44	[Baumgartner et al., 2005]
337	1974	F	+	+	+	–	MFS	Exon 10	c.1285C>T	p.R429X	[Mátyás et al., 2002a]
47	1963	F	+	(+)	+	–	MFS	Exon 12	c.1522delC	p.Q508R*fs*X71	This study
36	1961	B_P, F	+	–	+	–	MFS	Exon 12	c.1546C>T	p.R516X	[Arbustini et al., 2005]
36B	1995	B_P	+	–	(+)	+	MFS	Exon 12	c.1546C>T†	p.R516X†	[Arbustini et al., 2005]
372	1963	NA	+	–	+	+	MFS	Exon 13	c.1693C>T	p.R565X	[Quan et al., 1995]
55	1971	F	+	–	+	+	MFS	Exon 15	c.1904_1919del	p.Y635S*fs*X78	[Rommel et al., 2002]
326	1975	F	+	+	+	–	MFS	Exon 15	c.1904_1919del§	p.Y635S*fs*X78§	[Rommel et al., 2002]
305	1965	NA	+	(+)	+	+	MFS	Exon 15	c.1905C>G	p.Y635X	This study
262	1995	NA	(+)	–	+	–	MFS*	Exon 16	c.1981dupT	p.C661L*fs*X20	This study
233	1983	NA	+	–	+	+	MFS	Exon 17	c.2122C>T	p.Q708X	This study
443	2001	NA	+	–	–	–	MFS*	Exon 21	c.2581C>T	p.R861X	[Liu et al., 1997b]
249	1991	NA	+	?	+	?	MFS	Exon 23	c.2763T>A	p.C921X	This study
KF	1981	NA	(+)	–	+	+	MFS	Exon 25	c.3191_3194delAAGA	p.E1065G*fs*X22	[Baumgartner et al., 2005]
162B	1948	NA	–	–	+	+	MFS*	Exon 27	c.3373C>T	p.R1125X	[Rommel et al., 2005]
185	1957	F	?	?	?	?	MFS**	Exon 27	c.3373C>T§	p.R1125X§	[Rommel et al., 2005]
98	1956	F	–	+	+	?	MFS	Exon 28	c.3513C>A	p.C1171X	This study
61	1936	F	+	–	+	+	MFS	Exon 28	c.3568_3578del	p.P1190I*fs*X4	This study
13	1986	F	+	?	(+)	–	MFS*	Exon 35	c.4451delA	p.N1484T*fs*X5	This study
BP	1965	B_P, B_E, F	+	–	+	+	MFS	Exon 37	c.4621C>T	p.R1541X	[Halliday et al., 1999]
1	1975	B_P	+	–	(+)	?	MFS*	Exon 38	c.4786C>T	p.R1596X	[Loeys et al., 2001]
62	1961	NA	+	+	+	?	MFS	Exon 38	c.4786C>T§	p.R1596X§	[Loeys et al., 2001]
30	1987	NA	+	?	(+)	?	MFS*	Exon 41	c.5143delA	p.T1715P*fs*X178	This study
291	1976	B_P	+	–	(+)	?	MFS*	Exon 41	c.5157_5158insGG	p.C1720G*fs*X174	This study
63B	1979	NA	+	–	–	+	MFS*	Exon 43	c.5368C>T	p.R1790X	[Arbustini et al., 2005]
101	1960	B_P, F	+	–	+	?	MFS	Exon 43	c.5368C>T§	p.R1790X§	[Arbustini et al., 2005]
280	1983	B_P, F	(+)	–	+	–	MFS*	Exon 48	c.5924_5925dupAT	p.E1976M*fs*X5	This study
204	1987	F	?	?	?	?	MFS**	Intron 51	c.6379+6T>G (c.6379_6380insGTCAG)	p.D2127G*fs*X35	This study
56	1970	F	+	–	+	–	MFS*	Exon 57	c.7039_7040delAT	p.M2347V*fs*X19	[Körkkö et al., 2002]
281	1997	B_P	+	–	+	–	MFS	Exon 57	c.7063dupA	p.R2355K*fs*X12	This study
288	1966	B_P	+	–	+	+	MFS	Exon 58	c.7240C>T	p.R2414X	[Körkkö et al., 2002]
376	1969	B_P	(+)	+	(+)	+	MFS*	Exon 58	c.7240C>T§	p.R2414X§	[Körkkö et al., 2002]
139	1988	NA	+	–	+	–	MFS	Exon 61	c.7605C>A	p.C2535X	[Palz et al., 2000]
317	1993	NA	(+)	–	(+)	–	MFS*	Exon 62	c.7800C>A	p.Y2600X	This study
2B	1960	NA	+	+	+	?	MFS	Exon 62	c.7801C>T	p.Q2601X	[Baumgartner et al., 2005]
161	1963	NA	+	+	+	+	MFS	Exon 62	c.7801C>T§	p.Q2601X§	[Baumgartner et al., 2005]
344	1992	B_P	(+)	–	–	?	MFS*	Exon 63	c.7977C>A	p.C2659X	This study
455	1994	NA	+	–	+	?	MFS	Exon 65	c.8525_8529dcITTAAC	p.L2842P*fs*X7	[Palz et al., 2000]

[a]Coded designations (two-letter code or numbers).
[b]Material for transcript analysis: B_P, RNA extracted from PAXgene-stabilized whole blood; B_E, RNA extracted from EDTA-anticoagulated whole blood; F, RNA extracted from primary dermal fibroblasts; NA, material not available for transcript analysis.
[c]Patients indicated by MFS fulfilled the diagnostic criteria for Marfan syndrome [De Paepe et al., 1996], those by MFS* only partially. MFS** indicates patients with suspected Marfan syndrome without clinical details. SS, skeletal system; OS, ocular system; CS, cardiovascular system; FH, family history; +, affected; (+), slightly affected; –, not affected; ?, lack of information.
[d]Sequence variants are described in relation to the translation initiation site of the mRNA reference sequence NM_000138.3. Mutation numbering is after HGVS and journal standards, with +1 as the A of the ATG initiation codon. Recurrent mutations in related (†) and unrelated (§) patients are marked.

on a MJ Research PTC-200 thermocycler (MJ Research Inc., Waltham, MA; www.mjresearch.com).

Improvement of Quantitative Sequencing

In a preliminary study, we performed a series of tests to improve a previously described quantification of single nucleotide polymorphisms (SNPs) by automated DNA sequencing [Qiu et al., 2003]. In repeated (5×) experiments, we tested factors thought to be involved in the reproducibility and accuracy of quantitative sequencing. Experiments using gDNA samples of heterozygous individuals (C/T, C/A, G/A, and G/T) or gDNA pools with different wild-type to mutant allele frequencies (G/A, C/T) showed that (1) the number of PCR cycles (25, 32, 35, and 40) had, surprisingly, no impact on quantitative sequencing if (2) the amount of PCR templates was high enough to generate real-time PCR threshold cycle (*Ct*) values less than 30; (3) there was no significant difference between the DNA polymerases tested (HotStarTaq, Qiagen, Hilden, Germany; www.qiagen.com vs. HOT FIREPol, Solis Biodyne, Tartu, Estonia; www.sbd.ee); (4) for real-time PCR, 10 × SYBR Green I buffer (SYBR Green PCR Core Reagents, P/N 4304886, Applied Biosystems, Rotkreuz,

Switzerland; www.appliedbiosystems.com) was superior at 0.25 × rather than at 1 ×, 0.5 ×, or 0.125 × final concentration, also in comparison with 1 × ready to use SYBR Green PCR Master Mix (P/N 4309155, Applied Biosystems, Rotkreuz, Switzerland; www.appliedbiosystems.com); (5) ExoSAP-IT (USB Corporation, Cleveland, OH; www.usbweb.com) PCR purification resulted in more reproducible quantification results than column-based purification (QIAquick, Qiagen, Hilden, Germany; www.qiagen.com) or amplicons without purification; (6) sequences and positions of PCR and sequencing primers had no impact on quantitative sequencing if (7) peak heights in raw data chromatograms reached at least ~1,000 relative fluorescence units (RFU) but were below the detection maximum of ~8,000 RFU on an ABI3100 and if (8) raw data or the KB basecaller and corresponding mobility files were used instead of the original ABI basecaller and mobility files; (9) for the highest reproducibility the same capillary, polymer, and sequencing instrument should be used during the whole procedure of quantitative sequencing (POP6, POP7, ABI3730, and two different ABI3100 were tested). In addition, one-step RT-PCR experiments resulted in more reproducible transcript ratios than two-step protocols, also enabling the usage of the same reaction mix for both gDNA with known and RNA with unknown allele ratios.

Using the most reproducible conditions (Ct < 30, nonoverloaded raw data > 1,000 RFU, and KB basecaller or raw data), we tested the accuracy of our improved quantitative sequencing by analyzing (1) gDNA samples of patients with Charcot-Marie-Tooth disease type 1A (CMT1A) who carry an additional copy (duplication) of the *PMP22* gene and thus harbor the SNP c.*59A>C (NM_000304.2; rs13422 dbSNP build 129) in a natural ratio of 2:1 (66.7% vs. 33.3%); (2) four different gDNA pools with expected wild-type to mutant allele frequencies of 1.5:1, 2:1, 3:1, and 5:1 prepared by diluting the heterozygous mutant gDNA using homozygous wild-type gDNA as diluent (performed for two G>A substitutions in the *FBN1* exons 33 and 53); as well as (3) *FBN1* transcripts harboring single nucleotide substitutions (c.1875T>C, c.4621C>T) previously tested by our SNuPE-ONCE assay at cDNA level [Mátyás et al., 2002b].

Improved Quantitative Sequencing of Mutant Transcripts

Our quantitative sequencing approach consists of three main steps, namely, (1) determination of correction factors (for correction of unequal detection of the two alleles) that are derived from samples with known allele ratio, (2) peak height measurement of samples with unknown allele ratio, and (3) calculation of the relative amount of mutant allele using the corresponding correction factor.

Accordingly, for the determination of correction factors, gDNAs of patients heterozygous for an investigated mutation (A>B) were amplified in duplicate and each amplicon was sequenced four times in both directions, making sure that raw sequence signals are higher than 1,000 RFU but not overloaded. Peak heights (A_{50}, B_{50}) were obtained either from raw or analyzed data using Sequence Analysis Software v3.7/5.1/5.2 (Applied Biosystems, Rotkreuz, Switzerland; www.appliedbiosystems.com) and the corresponding KB basecaller v1.1/1.2. For raw sequence data, allele-specific peak heights were calculated by subtracting the lowest RFU (i.e., baseline signal) from the maximum RFU of the corresponding nucleotide base. For KB analyzed data, allele specific peak heights were determined by direct peak height measurements. Subsequently, correction factor (k) was calculated for each sequencing direction as an average of simple ratios of peak heights (A_{50}/B_{50}). Similarly, for the analysis of transcripts with unknown allele ratio, one-step RT-PCRs were performed in triplicate, each amplicon was sequenced four times in both directions, and peak heights (A_X, B_X) were obtained as measured at the gDNA level for known allele ratios. For the assessment of unknown allele ratios, mutant peak heights (B_X) were normalized by the corresponding correction factor ($k*B_X$) and the relative amount of the mutant allele (X_B) was estimated as an average of $X_B = k*B_X/(A_X + k*B_X)$ calculated for each sequencing direction. For the arithmetic mean of replicates ($n = 4$ for gDNA and $n = 6$ for cDNA), upper and lower confidence limits were calculated using critical values of paired t test distribution.

All RT-PCR thermal cycles were run on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Rotkreuz, Switzerland; www.appliedbiosystems.com) without 9600 emulation but with dissociation stage and using 10 × SYBR Green I buffer (SYBR Green PCR Core Reagents, P/N 4304886, Applied Biosystems, Rotkreuz, Switzerland; www.appliedbiosystems.com) at 0.25 × final concentration, to assure Ct values <30. RT-PCR purification, cycle sequencing reaction/purification, as well as RT-PCR product quality controls were performed as described above for qualitative transcript analyses.

Results

Reproducibility and Accuracy of Quantitative Sequencing

Quantitative sequencing resulted in highly reproducible relative allele frequencies (e.g., for heterozygotes 50.0 ± 1.4%, $P = 0.05$; Table 2) if (real-time) PCR generated Ct values <30, raw sequence signals were high enough but not overloaded, and, most importantly, KB basecaller or raw sequence data was used for peak height measurement. Indeed, in contrast to the original ABI basecaller, raw sequence data and the KB basecaller provided highly reproducible results, independent of the start and end

Table 2. Reproducibility and Accuracy of the Improved Quantitative Sequencing Assay Exemplified by Different Allele Ratios and Templates

Patient	Gene	Mutation	Template	Expected allele ratio (%)	Allele ratio (%) observed by quantitative sequencing (CI)
24108	*PMP22*	c.*59A>C (3′UTR)	gDNA	50.0:50.0[a]	50.0:50.0 (±1.4)
22564	*PMP22*	c.*59A>C (3′UTR)	gDNA	66.7:33.3[a]	66.6:33.4 (±2.2)
BP	*FBN1*	c.4621C>T (p.R1541X)	cDNA[b]	70.7:29.3[c]	71.8:28.2 (±2.7)[d]
RB	*FBN1*	c.1875T>C (p.N625N)	cDNA	50.3:49.7[c]	46.9:53.1 (±3.7)

[a]Patient 24108 is heterozygous for c.*59A>C and thus represents a natural 1:1 (50.0% vs. 50.0%) DNA pool of normal and mutant *PMP22* alleles (NM_000304.2; rs13422 dbSNP build 129). Patient 22564 is affected by the Charcot-Marie-Tooth disease type 1A (CMT1A) and thus carries an additional copy (duplication) of the *PMP22* gene, harboring c.*59A>C in a natural ratio of 2:1 (66.7% vs. 33.3%).
[b]Cells were treated with cycloheximide prior to RNA extraction.
[c]Allele ratios previously determined by using the SNuPE-ONCE assay [Mátyás et al., 2002b].
[d]Prior to quantitative sequencing, RT-PCR product was column purified using the QIAquick PCR purification kit as used also for the SNuPE-ONCE assay [Mátyás et al., 2002b].
CI, 95% confidence interval.

points used for sequence analysis. The lower detection limit of quantitative sequencing ranged from 3 to 9% depending on the heights of background signals (Fig. 1).

The quantitative sequence analyses of *PMP22* gene duplication and gDNA pools with four different wild-type to mutant allele ratios provided excellent accuracy (e.g., for *PMP22* duplication 66.6 ± 2.2%, $P = 0.05$; Table 2) and linear correlation (Pearson's correlation coefficient $R > 0.99$, data not shown) between expected and observed relative allele frequencies, respectively. Our SNuPE-ONCE and quantitative sequencing assays revealed comparable accuracies (Table 2). In contrast to SNuPE-ONCE, however, quantitative sequencing needs only primers and chemistries used for mutation detection in routine sequencing, making it superior for quantification of relative allele frequencies of > 10%.

Qualitative Transcript Analyses

Because not only intronic but also exonic mutations can affect *FBN1* pre-mRNA splicing, for example, by disrupting exonic splicing enhancer (ESE) sequences by the mechanism called nonsense-mediated alternative splicing (NAS) leading to exon skipping or aberrant splicing [cf. Caputi et al., 2002], we performed qualitative transcript analyses for all 26 patients for which cycloheximide-treated fibroblast cell lines and/or PAXgene-stabilized blood samples were available (Table 1). Qualitative sequencing and fragment length analysis of PTC-introducing exonic mutations revealed no aberrant splicing, although mutant transcripts were clearly detectable by sequencing. The three intronic nucleotide changes (Patients 59, 46, 46B, and 204) resulted in different splicing defects. Whereas c.247+1G>A (Patient 59) led to skipping of the entire exon 2, c.347−2A>G (Patients 46 and 46B) introduced a novel acceptor splice site and thus inserted one nucleotide (G) from intron 3 into the open reading frame (Table 1). Similarly, c.6379+6T>G (Patient 204) created a novel donor splice site, extending exon 51 by five nucleotides of intron 51 (Table 1). The splicing defects of c.247+1G>A and c.347−2A>G were identical both in cycloheximide-treated fibroblasts and PAXgene blood samples, while the effect of c.6379+6T>G could be analyzed only in fibroblasts. All three intronic mutations introduced new reading frames ending in a PTC.

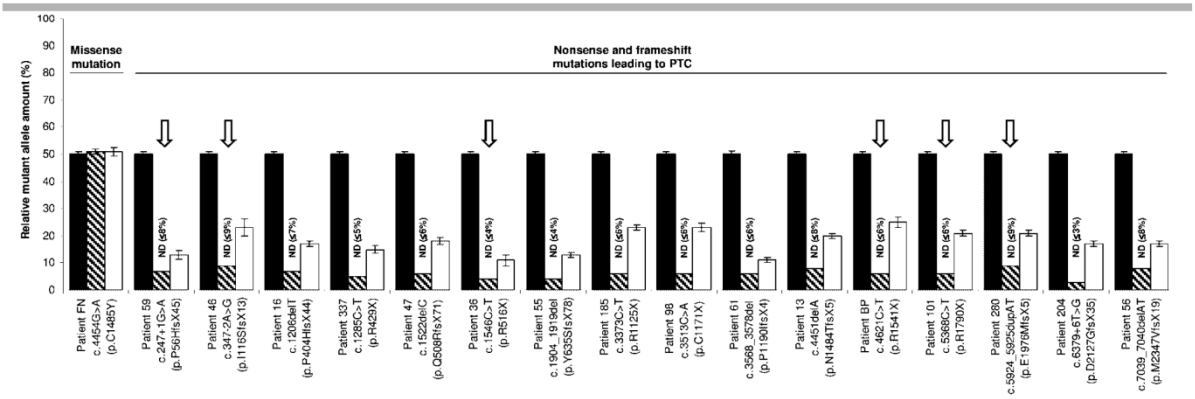

Figure 1. Quantitative sequence analysis of premature termination codon (PTC)-containing *FBN1* transcripts in fibroblasts. The relative amounts of mutant alleles are given with respect to the total heights of corresponding electrophoretic peaks normalized by the appropriate correction factor derived from gDNA (black bars). Mutant transcripts from cycloheximide-untreated (bars with diagonal lines) and cycloheximide-treated (white bars) fibroblasts are indicated. Note that in all untreated fibroblasts the relative amounts of mutant alleles were below the detection limit of the assay (ND, not detectable, the highest level of background signals are given in parentheses). The *FBN1* missense mutation c.4454G>A (p.C1485Y) was used as control. Error bars indicate 95% confidence intervals. Arrows denote patients from whom PAXgene-stabilized blood sample was also taken for further investigations (see Fig. 3B).

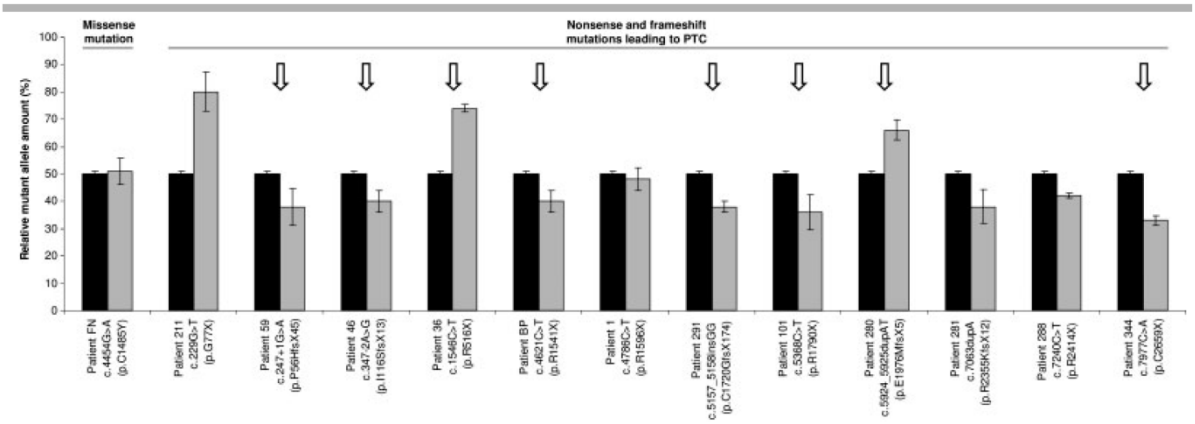

Figure 2. Quantitative sequence analysis of PTC-containing *FBN1* transcripts in PAXgene-stabilized blood samples, showing high levels of mutant transcripts (light gray bars). Arrows denote patients from whom a second PAXgene-stabilized blood sample was investigated (see Fig. 3A); other labels are as used in Figure 1.

Quantitative Analysis of PTC-containing Transcripts

In cycloheximide-untreated fibroblasts, quantitative sequencing revealed no mutant transcripts, that is, mutant alleles were below the detection limit of the assay (Fig. 1). In cycloheximide-treated fibroblasts, however, quantitative sequencing clearly detected PTC-containing transcripts with relative frequencies ranging, likely according to treatment efficiencies and cell responses, from $11.0 \pm 2.1\%$ to $25.0 \pm 1.8\%$ ($P = 0.05$; Fig. 1). As expected, cycloheximide treatment did not significantly change the level of transcripts harboring the missense mutation c.4454G>A (p.C1485Y) used as control in this study (Fig. 1). These results highly suggest the translation-dependent degradation of PTC-containing *FBN1* transcripts in fibroblasts.

In contrast, in PAXgene-stabilized blood samples quantitative sequencing demonstrated a high relative amount of PTC-containing *FBN1* transcripts ranging from $33.0 \pm 3.9\%$ to $80.0 \pm 7.2\%$ ($P = 0.05$), suggesting that NMD is incomplete in leukocytes (Figs. 2, 3B, and 4). Intraindividual (Figs. 3A and 4), interindividual (Fig. 3C), and intrafamilial (Fig. 3C) differences in the corresponding relative mutant transcript levels were statistically not significant in the majority of cases, indicating the impact of the quantitative

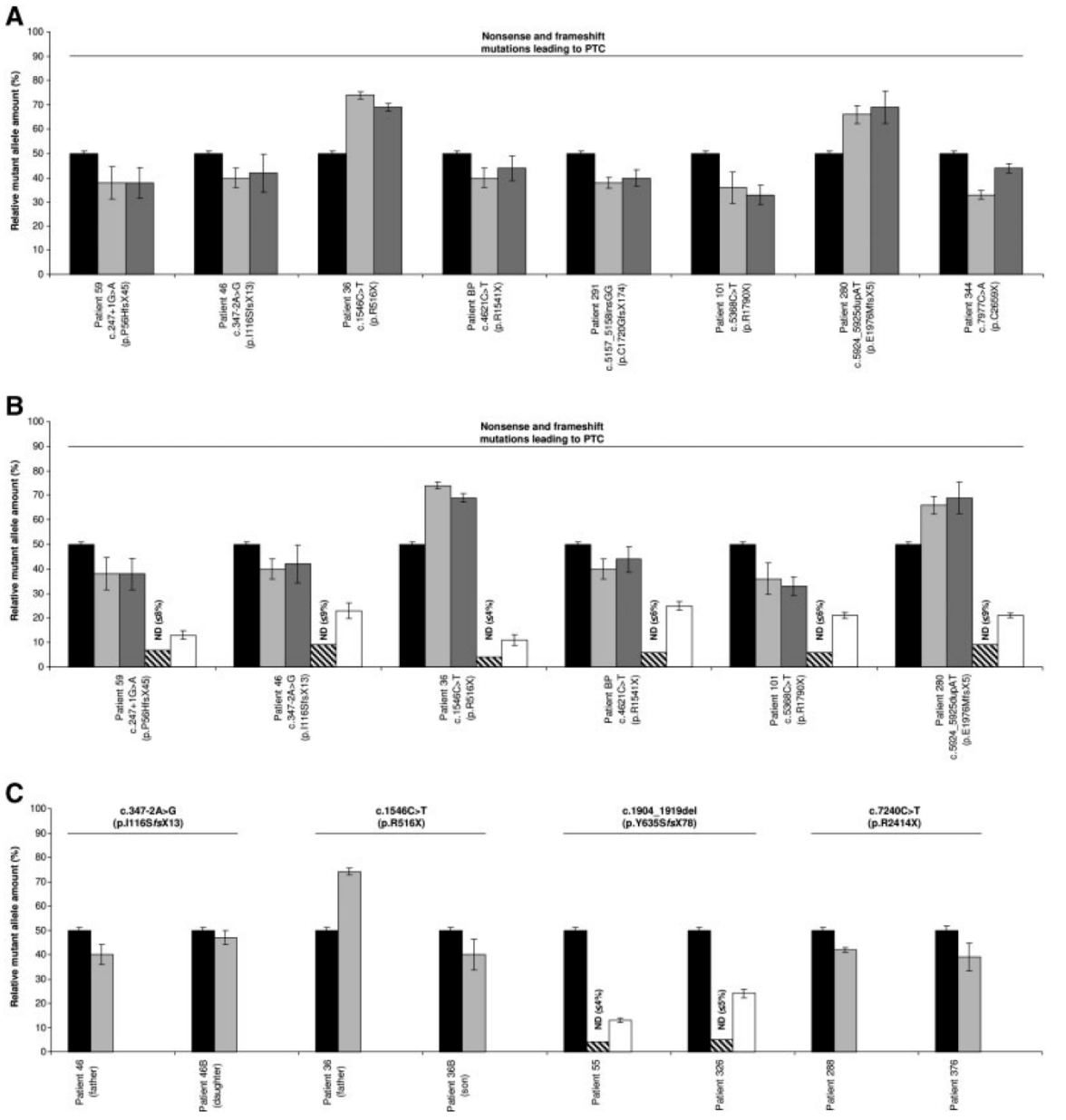

Figure 3. Variation in mutant transcript levels among different probes, tissues, and individuals. **A:** Quantitative sequence analysis of PTC-containing *FBN1* transcripts in two sequentially taken PAXgene-stabilized blood samples. **B:** Transcript levels in two sequentially taken PAXgene blood samples and fibroblasts. **C:** Transcripts of the same mutation in related (Patients 46 and 46B, 36, and 36B) and unrelated (Patients 55, 288, 326, and 376) individuals. Transcript levels of the second PAXgene blood samples are indicated by dark gray bars; other labels are as used in Figures 1 and 2.

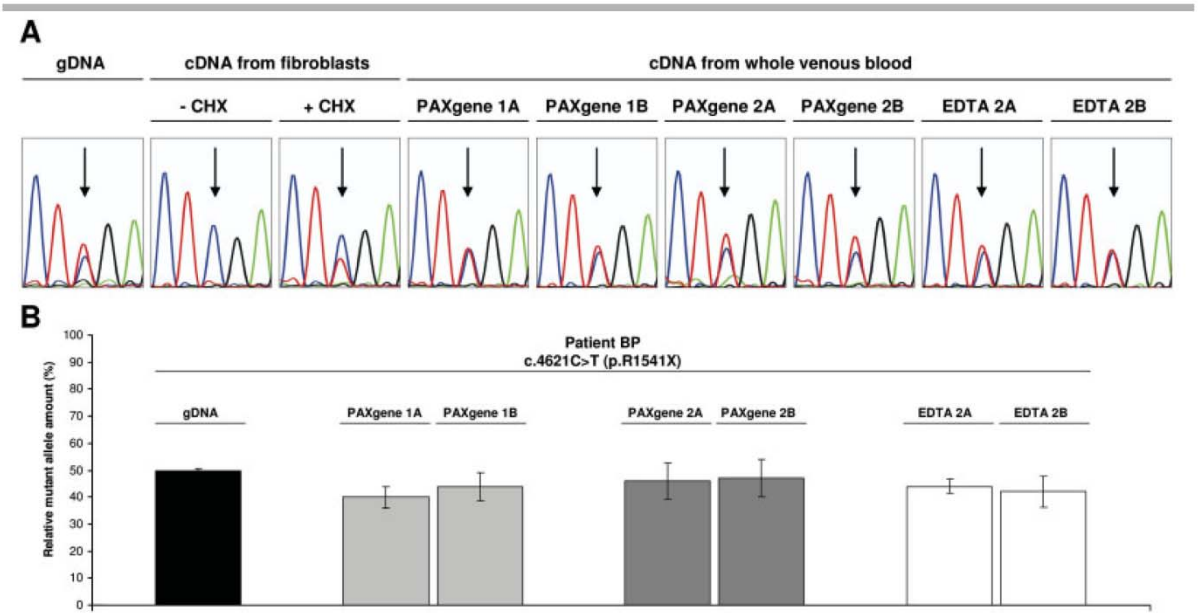

Figure 4. Quantitative sequencing exemplified by analysis of the *FBN1* nonsense mutation c.4621C>T (p.R1541X). **A:** Representative partial forward sequence of the region flanking c.4621C>T (arrowed). Sequences on gDNA with known relative mutant allele frequency of 50% were used to calculate the correction factor (*k*), which is the simple ratio of peak heights of the wild-type C (blue) and mutant T (red) bases, allowing the determination of unknown allele ratios at cDNA level. Fibroblasts were treated (+CHX) or untreated (−CHX) by cycloheximide prior to RNA extraction. Leukocyte RNA was extracted from four PAXgene-stabilized blood samples (PAXgene 1A and 1B taken in 2004, PAXgene 2A and 2B taken in 2007) and two fresh EDTA blood samples (EDTA 2A and 2B taken in 2007). **B:** Results of repeated quantitative sequencing of cDNAs derived from blood samples as exemplified in **A**. Note that different blood samples and RNA extraction methods result in no significant difference in the relative amounts of mutant transcript. Error bars indicate 95% confidence intervals. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

transcript analysis of a single PAXgene-stabilized blood sample. DNase treatment did not significantly influence the results of quantitative sequencing (data not shown). Similarly, comparative analysis of PAXgene and fresh EDTA blood samples demonstrated comparable relative amounts of PTC-bearing transcripts ranging from 40.0±4.0% to 47.0±7.1% ($P = 0.05$; Figs. 4A and B), suggesting that the stabilizing solution used in the PAXgene tubes does not interfere with the NMD machinery and thus does not influence the interpretation of our results.

Discussion

In the postgenome era, not only the detection and scoring of SNPs is important but also their quantification at both the gDNA and cDNA levels. Available technologies for SNP quantification, such as real-time PCR, pyrosequencing, single-base extension assays, MLPA, microarrays, and massive parallel sequencing, however, have variable sensitivity and may need special equipment, chemistry, and/or assay design [e.g., Goossens et al., 2009]. For this reason, Sanger DNA sequencing using dye terminator chemistry on standard sequencing machines has been suggested as a fast and cost-effective single-tube method for SNP quantification, taking advantage of the fact that the peak height of a particular nucleotide base is proportional to the template amount present in the reaction [e.g., Anjos et al., 2004; Lewin et al., 2004; Nurpeisov et al., 2003; Qiu et al., 2003]. However, one major problem with nonreal-time-PCR-based assays such as sequencing is to choose the number of thermal cycles and the amount of templates in PCR suitable for quantification. Our method described here can overcome this limitation by a single real-time amplification using SYBR Green I, allowing the characterization of PCR prior to sequencing. Using correction factors, KB basecaller, and identical conditions for both the determination of correction factors and the measurement of unknown allele ratios, our approach circumvents further drawbacks of previous studies, enabling reproducible and accurate quantification of SNP variants. Consequently, our method can be used not only for transcript analyses, as demonstrated here, but also for quantitative analysis of copy number variations, proportion of methylation, mosaicisms, and DNA pools.

Using our improved sequencing assay for quantitative analysis of a large cohort of PTC-introducing *FBN1* mutations, we found high levels of mutant transcripts in leukocytes, whereas mutant transcripts were undetectable in untreated fibroblasts. Although we have not investigated NMD directly, this finding highly suggests tissue-specific degradation of PTC-containing *FBN1* transcripts by NMD. This conclusion is in accordance with a previous study which showed that NMD of *COL10A1* nonsense mutations causing Schmid-type metaphyseal chondrodysplasia is complete in cartilage chondrocytes but incomplete in noncartilage lymphoblastoid cell lines and bone-derived osteoblast cultures [Bateman et al., 2003]. A difference in NMD has also been previously observed between blood samples and lymphoblastoid cell lines of the same patients, demonstrating high levels of PTC-bearing *OPA1* transcripts in RNA obtained from blood (40–50%) in comparison with that from lymphoblastoid cell lines (20–30%) [Schimpf et al., 2008]. In addition, intertissue (CVS-stromal cells, amniocytes, and lymphocytes) and interindividual variation in NMD efficiency has been reported for PTCs in different genes, such as *DMD* and *ESCO2* [Kerr et al., 2001; Resta et al., 2006]. Further evidence has recently been provided for variable NMD

Table 3. Expression of PTC-Containing Transcripts of Different Genes Previously Determined in PAXgene-Stabilized Blood

Gene			PTC		
Symbol	Name	Exons	Position	Expression (based on)	Reference
DSG2	Desmoglein 2	15	Exon 13	~30–40% (agarose gel)	[Pilichou et al., 2006]
OPA1	Optic atrophy 1	30	Exons 11, 15, 20, 21, 25, 27	20–40% (pyrosequencing)	[Schimpf et al., 2008]
MLH1	MutL homolog 1, colon cancer, nonpolyposis type 2	19	Exon 14	<10% (agarose gel)	[Wehner et al., 2005]
DFNB59	Deafness, autosomal recessive 59	7	Exon 4	~30% (agarose gel)	[Collin et al., 2007]
LMNA	Lamin A	12	Exon 8	~20–30% (agarose gel)	[Morel et al., 2006]
SGCE	Sarcoglycan, epsilon	11	Exon 6	~40–50% (agarose gel)	[Misbahuddin et al., 2007]

efficiency among cells, suggesting that NMD is an inherent character of each cell [Linde et al., 2007].

One explanation for this variability in the efficiency of NMD may be tissue- and/or cell type-specific expression and interaction of NMD factors, such as UPF1, UPF2, and UPF3, and components of the EJC, such as eIF4A3, Y14, MAGOH, and RNPS1. Indeed, different NMD routes have been specified by distinct EJC components, in which low abundance of RNPS1 has been associated with low NMD efficiency [e.g., Chan et al., 2007; Gehring et al., 2005; Viegas et al., 2007]. Thus, it might be possible, for example, that the expression of the *RNPS1* gene is more abundant in fibroblasts leading to intensive degradation of PTC-bearing *FBN1* transcripts. Furthermore, the presence of different NMD routes in blood and fibroblasts cannot be excluded due to lack of reports on tissue-specific variation of NMD routes. Alternatively, it is possible that *FBN1* transcripts are present at such a low level that it is under the detection limit of the NMD machinery. This hypothesis is consistent with our observation that the total amount of *FBN1* transcripts in blood is ~1,000 times lower than that expressed in fibroblasts (data not shown). Similarly, because the 5′ end of the *FBN1* gene consists of three alternatively spliced exons as well as sequences with different promoter activity [Biery et al., 1999; Corson et al., 1993; Guo et al., 2007], one could speculate that *FBN1* transcripts in blood use other regulatory sequences generating an alternative open reading frame resistant to NMD. However, we could not explain incomplete NMD in blood by in silico six-frame translation of the entire known coding sequence of *FBN1* (data not shown). In addition, the variation in the relative amounts of mutant alleles observed in this study may not only be due to variation in NMD but also to differences in the expression of both mutant and normal *FBN1* alleles [cf. Hutchinson et al., 2003]. Further studies are necessary to clarify these possibilities.

Moreover, in this study we also extended the number of disease-causing *FBN1* mutations by identifying 18 novel PTC-introducing sequence variants. By demonstrating the true effect of PTC mutations in fibroblasts that trigger the NMD quality control pathway, the mutations identified here can serve as a basis for genotype-phenotype studies [cf. Faivre et al., 2007, 2009], emphasizing the role of functional haploinsufficiency in the molecular pathogenesis of MFS [cf. Mátyás et al., 2007]. Although several studies have previously investigated the expression of *FBN1* transcripts in fibroblasts [e.g., Dietz et al., 1993a,b; Halliday et al., 1999; Hewett et al., 1994; Schrijver et al., 2002], this work describes the first comparative transcript analysis of PTC-introducing *FBN1* mutations in whole blood and fibroblast cells. Our finding that PTC-bearing transcripts are clearly detectable in PAXgene blood samples seems to be neither restricted to *FBN1* nor universally valid (Table 3). This may also be true for lymphoblastoid cell lines, albeit with lower PTC transcript levels as discussed above, because incomplete NMD has also been observed in such cells investigating PTC-containing transcripts of the genes *OPA1*, *BRCA1*, and *APC* [Gismondi et al., 1998; Perrin-Vidoz et al., 2002; Schimpf et al., 2008]. However, it remains unknown whether incomplete NMD occurs with comparable levels of PTC transcripts in all types of leukocytes.

Finally, incomplete NMD has consequences for molecular genetic testing. Incomplete NMD can allow RNA-based mutation detection in blood, offering efficient screening of the exons of large genes as well as identification of aberrant splicing caused by intronic sequence variants undetectable by standard exon-by-exon gDNA screening. However, routine RNA-based mutation screening in blood may be hampered by low gene expression and RNA quality (RIN 5–7) as well as by relatively short RT-PCR product lengths (400–500 bp), as is the case for *FBN1*. In addition, while PTC mutations and splicing defects may be clearly detected, no correct conclusion is possible on mutant transcript expression and NMD in other (affected) tissues by analyzing RNA from blood samples. Furthermore, for some genes (e.g., *COL3A1*) RT-PCR may fail or results in unspecific products both in PAXgene blood and lymphoblastoid cells (data not shown), making transcript analyses in such samples difficult or impossible. Consequently, the a priori possibility of incomplete NMD and/or mutation detection should be considered in transcript analyses for each gene, PTC mutation, and tissue.

Acknowledgments

We are grateful to the patients for participating in this study and the referring physicians (Klaus Ammann, Deborah Bartholdi, Sophie Dahoun, Alexandre Dayer, Nursel Elcioglu, Claudio Fabris, Bruno Frischkopf, Monica Gersbach-Forrer, Gerber Glanzmann, Oswald Hasselmann, Karl Heinimann, Peter Hoessly, Daniela Kaiser, Thomas Lüscher, Hansjakob Müller, Kirsten Rasmussen, Albert Schinzel, Roger Simon, Barbara Utermann, and Hari Zvizdic) for help with obtaining appropriate samples and/or information about the clinical phenotypes of patients. We thank Angelika Schwarze for cell cultures; Melanie Maudrich for initial *FBN1* mutation analysis; Janine Meienberg for *COL3A1* transcript analyses; Silke Feil and Philippe Reuge for technical support; Helen Burri, Caroline Henggeler, Barbara Kloeckener, Regina Perez, Elisabeth Probst, Nikolaus Schäfer, and other members of the Institute of Medical Genetics, University of Zurich, for constructive discussions. We also thank Burkhardt Seifert and Kaspar Rufibach for statistical discussions and suggestions. This work was supported by grants from the Foundation for Research at the Medical Faculty and Research Funding of the University of Zurich (to I.M. and G.M.), Swiss Heart Foundation (to G.M.), Swiss National Science Foundation (NF 3200B0-109370/1 to B.S.; NF 3100A0-120504 to G.M.), and the Wolfermann-Nägeli-Stiftung Zurich (to B.S.).

References

Andreutti-Zaugg C, Scott RJ, Iggo R. 1997. Inhibition of nonsense-mediated messenger RNA decay in clinical samples facilitates detection of human MSH2 mutations with an in vivo fusion protein assay and conventional techniques. Cancer Res 57:3288–3293.

Anjos SM, Tessier MC, Polychronakos C. 2004. Association of the cytotoxic T lymphocyte-associated antigen 4 gene with type 1 diabetes: evidence for independent effects of two polymorphisms on the same haplotype block. J Clin Endocrinol Metab 89:6257–6265.

Arbustini E, Grasso M, Ansaldi S, Malattia C, Pilotto A, Porcu E, Disabella E, Marziliano N, Pisani A, Lanzarini L, Mannarino S, Larizza D, Mosconi M, Antoniazzi E, Zoia MC, Meloni G, Magrassi L, Brega A, Bedeschi MF, Torrente I, Mari F, Tavazzi L. 2005. Identification of sixty-two novel and twelve known *FBN1* mutations in eighty-one unrelated probands with Marfan syndrome and other fibrillinopathies. Hum Mutat 26:494.

Bateman JF, Freddi S, Nattrass G, Savarirayan R. 2003. Tissue-specific RNA surveillance? Nonsense-mediated mRNA decay causes collagen X haploinsufficiency in Schmid metaphyseal chondrodysplasia cartilage. Hum Mol Genet 12:217–225.

Baumgartner D, Baumgartner C, Mátyás G, Steinmann B, Löffler-Ragg J, Schermer E, Schweigmann U, Baldissera I, Frischhut B, Hess J, Hammerer I. 2005. Diagnostic power of aortic elastic properties in young patients with Marfan syndrome. J Thorac Cardiovasc Surg 129:730–739.

Beroud C, Collod-Beroud G, Boileau C, Soussi T, Junien C. 2000. UMD (Universal mutation database): a generic software to build and analyze locus-specific databases. Hum Mutat 15:86–94.

Biery NJ, Eldadah ZA, Moore CS, Stetten G, Spencer F, Dietz HC. 1999. Revised genomic organization of FBN1 and significance for regulated gene expression. Genomics 56:70–77.

Caputi M, Kendzior Jr RJ, Beemon KL. 2002. A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer. Genes Dev 16:1754–1759.

Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285–298.

Chan WK, Huang L, Gudikote JP, Chang YF, Imam JS, MacLean 2nd JA, Wilkinson MF. 2007. An alternative branch of the nonsense-mediated decay pathway. EMBO J 26:1820–1830.

Chang YF, Imam JS, Wilkinson MF. 2007. The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76:51–74.

Collin RW, Kalay E, Oostrik J, Caylan R, Wollnik B, Arslan S, den Hollander AI, Birinci Y, Lichtner P, Strom TM, Toraman B, Hoefsloot LH, Cremers CW, Brunner HG, Cremers FP, Karaguzel A, Kremer H. 2007. Involvement of DFNB59 mutations in autosomal recessive nonsyndromic hearing impairment. Hum Mutat 28:718–723.

Corson GM, Chalberg SC, Dietz HC, Charbonneau NL, Sakai LY. 1993. Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure and alternatively spliced exons at the 5′ end. Genomics 17:476–484.

De Paepe A, Devereux RB, Dietz HC, Hennekam RC, Pyeritz RE. 1996. Revised diagnostic criteria for the Marfan syndrome. Am J Med Genet 62:417–426.

Dietz HC, Cutting GR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, Puffenberger EG, Hamosh A, Nanthakumar EJ, Curristin SM, Stetten G, Meyers DA, Francomano CA. 1991. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. Nature 352:337–339.

Dietz HC, McIntosh I, Sakai LY, Corson GM, Chalberg SC, Pyeritz RE, Francomano CA. 1993a. Four novel FBN1 mutations: significance for mutant transcript level and EGF-like domain calcium binding in the pathogenesis of Marfan syndrome. Genomics 17:468–475.

Dietz HC, Valle D, Francomano CA, Kendzior Jr RJ, Pyeritz RE, Cutting GR. 1993b. The skipping of constitutive exons in vivo induced by nonsense mutations. Science 259:680–683.

Faivre L, Collod-Beroud G, Callewaert B, Child A, Binquet C, Gautier E, Loeys BL, Arbustini E, Mayer K, Arslan-Kirchner M, Stheneur C, Kiotsekoglou A, Comeglio P, Marziliano N, Wolf JE, Bouchot O, Khau-Van-Kien P, Beroud C, Claustres M, Bonithon-Kopp C, Robinson PN, Adès L, De Backer J, Coucke P, Francke U, De Paepe A, Jondeau G, Boileau C. 2009. Clinical and mutation-type analysis from an international series of 198 probands with a pathogenic FBN1 exons 24–32 mutation. Eur J Hum Genet 17:491–501.

Faivre L, Collod-Beroud G, Loeys BL, Child A, Binquet C, Gautier E, Callewaert B, Arbustini E, Mayer K, Arslan-Kirchner M, Kiotsekoglou A, Comeglio P, Marziliano N, Dietz HC, Halliday D, Beroud C, Bonithon-Kopp C, Claustres M, Muti C, Plauchu H, Robinson PN, Adès LC, Biggin A, Benetts B, Brett M, Holman KJ, De Backer J, Coucke P, Francke U, De Paepe A, Jondeau G, Boileau C. 2007. Effect of mutation type and location on clinical outcome in 1,013 probands with Marfan syndrome or related phenotypes and FBN1 mutations: an international study. Am J Hum Genet 81:454–466.

Frischmeyer PA, Dietz HC. 1999. Nonsense-mediated mRNA decay in health and disease. Hum Mol Genet 8:1893–1900.

Gehring NH, Kunz JB, Neu-Yilik G, Breit S, Viegas MH, Hentze MW, Kulozik AE. 2005. Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. Mol Cell 20:65–75.

Gismondi V, Stagnaro P, Pedemonte S, Biticchi R, Presciuttini S, Grammatico P, Sala P, Bertario L, Groden J, Varesco L. 1998. Chain-terminating mutations in the APC gene lead to alterations in APC RNA and protein concentration. Genes Chromosomes Cancer 22:278–286.

Goossens D, Moens LN, Nelis E, Lenaerts AS, Glassee W, Kalbe A, Frey B, Kopal G, Jonghe PD, Rijk PD, Del-Favero J. 2009. Simultaneous mutation and copy number variation (CNV) detection by multiplex PCR-based GS-FLX sequencing. Hum Mutat 30:472–476.

Guo G, Bauer S, Hecht J, Schulz MH, Busche A, Robinson PN. 2007. A short ultraconserved sequence drives transcription from an alternate FBN1 promoter. Int J Biochem Cell Biol 40:638–650.

Halliday D, Hutchinson S, Kettle S, Firth H, Wordsworth P, Handford PA. 1999. Molecular analysis of eight mutations in FBN1. Hum Genet 105:587–597.

Hewett D, Lynch J, Child A, Firth H, Sykes B. 1994. Differential allelic expression of a fibrillin gene (FBN1) in patients with Marfan syndrome. Am J Hum Genet 55:447–452.

Hutchinson S, Furger A, Halliday D, Judge DP, Jefferson A, Dietz HC, Firth H, Handford PA. 2003. Allelic variation in normal human FBN1 expression in a family with Marfan syndrome: a potential modifier of phenotype? Hum Mol Genet 12:2269–2276.

Inacio A, Silva AL, Pinto J, Ji X, Morgado A, Almeida F, Faustino P, Lavinha J, Liebhaber SA, Romao L. 2004. Nonsense mutations in close proximity to the initiation codon fail to trigger full nonsense-mediated mRNA decay. J Biol Chem 279:32170–32180.

Karttunen L, Ukkonen T, Kainulainen K, Syvänen AC, Peltonen L. 1998. Two novel fibrillin-1 mutations resulting in premature termination codons but in different mutant transcript levels and clinical phenotypes. Hum Mutat Suppl 1:S34–S37.

Kerr TP, Sewry CA, Robb SA, Roberts RG. 2001. Long mutant dystrophins and variable phenotypes: evasion of nonsense-mediated decay? Hum Genet 109: 402–407.

Khajavi M, Inoue K, Lupski JR. 2006. Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. Eur J Hum Genet 14:1074–1081.

Körkkö J, Kaitila I, Lönnqvist L, Peltonen L, Ala-Kokko L. 2002. Sensitivity of conformation sensitive gel electrophoresis in detecting mutations in Marfan syndrome and related conditions. J Med Genet 39:34–41.

Lewin J, Schmitt AO, Adorján P, Hildmann T, Piepenbrock C. 2004. Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplificates. Bioinformatics 20:3005–3012.

Linde L, Boelz S, Neu-Yilik G, Kulozik AE, Kerem B. 2007. The efficiency of nonsense-mediated mRNA decay is an inherent character and varies among different cells. Eur J Hum Genet 15:1156–1162.

Liu W, Qian C, Francke U. 1997a. Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. Nat Genet 16:328–329.

Liu WO, Oefner PJ, Qian C, Odom RS, Francke U. 1997b. Denaturing HPLC-identified novel FBN1 mutations, polymorphisms, and sequence variants in Marfan syndrome and related connective tissue disorders. Genet Test 1:237–242.

Loeys B, Nuytinck L, Delvaux I, De Bie S, De Paepe A. 2001. Genotype and phenotype analysis of 171 patients referred for molecular study of the fibrillin-1 gene FBN1 because of suspected Marfan syndrome. Arch Intern Med 161:2447–2454.

Maquat LE. 1995. When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. RNA 1:453–465.

Mátyás G, Alonso S, Patrignani A, Marti M, Arnold E, Magyar I, Henggeler C, Carrel T, Steinmann B, Berger W. 2007. Large genomic fibrillin-1 (FBN1) gene deletions provide evidence for true haploinsufficiency in Marfan syndrome. Hum Genet 122:23–32.

Mátyás G, De Paepe A, Halliday D, Boileau C, Pals G, Steinmann B. 2002a. Evaluation and application of denaturing HPLC for mutation detection in Marfan syndrome: Identification of 20 novel mutations and two novel polymorphisms in the FBN1 gene. Hum Mutat 19:443–456.

Mátyás G, Giunta C, Steinmann B, Hossle JP, Hellwig R. 2002b. Quantification of single nucleotide polymorphisms: a novel method that combines primer extension assay and capillary electrophoresis. Hum Mutat 19:58–68.

Mendell JT, Dietz HC. 2001. When the message goes awry: disease-producing mutations that influence mRNA content and performance. Cell 107:411–414.

Misbahuddin A, Placzek M, Lennox G, Taanman JW, Warner TT. 2007. Myoclonus-dystonia syndrome with severe depression is caused by an exon-skipping mutation in the epsilon–sarcoglycan gene. Mov Disord 22:1173–1175.

Morel CF, Thomas MA, Cao H, O'Neil CH, Pickering JG, Foulkes WD, Hegele RA. 2006. A LMNA splicing mutation in two sisters with severe Dunnigan-type familial partial lipodystrophy type 2. J Clin Endocrinol Metab 91:2689–2695.

Mort M, Ivanov D, Cooper DN, Chuzhanova NA. 2008. A meta-analysis of nonsense mutations causing human genetic disease. Hum Mutat 29:1037–1047.

Mühlemann O, Eberle AB, Stalder L, Zamudio Orozco R. 2008. Recognition and elimination of nonsense mRNA. Biochim Biophys Acta 1779:538–549.

Noensie EN, Dietz HC. 2001. A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. Nat Biotechnol 19:434–439.

Nurpeisov V, Hurwitz SJ, Sharma PL. 2003. Fluorescent dye terminator sequencing methods for quantitative determination of replication fitness of human immunodeficiency virus type 1 containing the codon 74 and 184 mutations in reverse transcriptase. J Clin Microbiol 41:3306–3311.

Palz M, Tiecke F, Booms P, Goldner B, Rosenberg T, Fuchs J, Skovby F, Schumacher H, Kaufmann UC, von Kodolitsch Y, Nienaber CA, Leitner C, Katzke S, Vetter B, Hagemeier C, Robinson PN. 2000. Clustering of mutations associated with mild Marfan-like phenotypes in the 3′ region of FBN1 suggests a potential genotype-phenotype correlation. Am J Med Genet 91:212–221.

Perrin-Vidoz L, Sinilnikova OM, Stoppa-Lyonnet D, Lenoir GM, Mazoyer S. 2002. The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons. Hum Mol Genet 11:2805–2814.

Pilichou K, Nava A, Basso C, Beffagna G, Bauce B, Lorenzon A, Frigo G, Vettori A, Valente M, Towbin J, Thiene G, Danieli GA, Rampazzo A. 2006. Mutations in desmoglein-2 gene are associated with arrhythmogenic right ventricular cardiomyopathy. Circulation 113:1171–1179.

Qiu P, Soder GJ, Sanfiorenzo VJ, Wang L, Greene JR, Fritz MA, Cai XY. 2003. Quantification of single nucleotide polymorphisms by automated DNA sequencing. Biochem Biophys Res Commun 309:331–338.

Quan F, Sakai L, Popovich BW. 1995. The identification of fibrillin mutations in Marfan syndrome using heteroduplex analysis and nucleotide sequencing. Am J Hum Genet 57(Suppl.):A333.

Resta N, Susca FC, Di Giacomo MC, Stella A, Bukvic N, Bagnulo R, Simone C, Guanti G. 2006. A homozygous frameshift mutation in the ESCO2 gene: evidence of intertissue and interindividual variation in Nmd efficiency. J Cell Physiol 209:67–73.

Rommel K, Karck M, Haverich A, Schmidtke J, Arslan-Kirchner M. 2002. Mutation screening of the fibrillin-1 (FBN1) gene in 76 unrelated patients with Marfan syndrome or Marfanoid features leads to the identification of 11 novel and three previously reported mutations. Hum Mutat 20:406–407.

Rommel K, Karck M, Haverich A, von Kodolitsch Y, Rybczynski M, Muller G, Singh KK, Schmidtke J, Arslan-Kirchner M. 2005. Identification of 29 novel and nine recurrent fibrillin-1 (*FBN1*) mutations and genotype-phenotype correlations in 76 patients with Marfan syndrome. Hum Mutat 26: 529–539.

Sanchez-Sanchez F, Ramirez-Castillejo C, Weekes DB, Beneyto M, Prieto F, Najera C, Mittnacht S. 2007. Attenuation of disease phenotype through alternative translation initiation in low-penetrance retinoblastoma. Hum Mutat 28:159–167.

Schimpf S, Fuhrmann N, Schaich S, Wissinger B. 2008. Comprehensive cDNA study and quantitative transcript analysis of mutant OPA1 transcripts containing premature termination codons. Hum Mutat 29:106–112.

Schrijver I, Liu W, Odom R, Brenn T, Oefner P, Furthmayr H, Francke U. 2002. Premature termination mutations in FBN1: distinct effects on differential allelic expression and on protein and clinical phenotypes. Am J Hum Genet 71:223–237.

Stalder L, Mühlemann O. 2008. The meaning of nonsense. Trends Cell Biol 18:315-321.

Viegas MH, Gehring NH, Breit S, Hentze MW, Kulozik AE. 2007. The abundance of RNPS1, a protein component of the exon junction complex, can determine the variability in efficiency of the Nonsense Mediated Decay pathway. Nucleic Acids Res 35:4542–4551.

Wehner M, Mangold E, Sengteller M, Friedrichs N, Aretz S, Friedl W, Propping P, Pagenstecher C. 2005. Hereditary nonpolyposis colorectal cancer: pitfalls in deletion screening in MSH2 and MLH1 genes. Eur J Hum Genet 13:983–986.

Wilkinson MF. 2005. A new function for nonsense-mediated mRNA-decay factors. Trends Genet 21:143–148.

Magyar et al., *Human Mutation*

Supp. Table S1. RT-PCR primers used in this study

#	Primer name	Primer sequence (5'-3')	Primer location (exon)	Total size of introns between primers (bp)	Product length (bp)	Annealing temperature (°C)
1	cFBN1_N87_F cFBN1_N323_R	GGA GGC TGG GAA CGT GAA G ATC TGG AGC CAC AGG AAG GAG	1 3	33,696	257	64
2	cFBN1_N361_F cFBN1_N584_R	CCT GTG GGG ATG GAT TTT GTT C GCC ACA CAC CTT CCT CCA TTG	3 5	14,253	223	62
3	cFBN1_N1096_F cFBN1_N1530_R	GGT CTC CAG GGG TCA CTG TC CGA ACC CTG GTT GTT AAT ACA CTC A	9 12	6,669	434	62
4	cFBN1_N1386_F cFBN1_N1666_R	GGA CGC TGC ATT CCA ACT C CGC AAT GAA AAC TGC CAT CT	11 13	5,097	280	58
5	cFBN1_N1945_F cFBN1_N4_R	GCT GGC ATC AGA TGG ACG TTA T GGC GCA ACA GCA TTC AGA TT	14 16	4,519	250	62
6	cFBN1_9_F cFBN1_10_R	GCG GCT TTG CTC TTG ATT CTG A CCG CCG CTT CTG TCC AGT TC	25 35	16,483	1279	58
7	cFBN1_N3451_F cFBN1_N3679_R	CCG CGT GTA TCG ACA TCA A ATC CCG GCT GAC AGC TAC AT	27/28 29	1,578	228	58
8	cFBN1_N4318_F cFBN1_N4548_R	ACG GGA AAG CCT GTG AAG ATA ATC AGG TGG GCA GTC ACA GA	34/35 36	2,099	230	58
9	cFBN1_N4533_F cFBN1_N4779_R	GAC TGC CCA CCT GAT TTT GAA TTC CCC TCC AGG ACA AAG AAT	36 38	2,388	246	58
10	cFBN1_N4682_F cFBN1_N4987_R	CTG CTG CTG TTC TCT GGG TAA A ATG TCC CTG GAC CAC AGA TTC	37 40	3,721	305	58
11	cFBN1_N5045_F cFBN1_N5318_R	TGG GGG AAA TAA TTG CAT GG ATC TCC CGG CAC TCA TCA ATA	40 43	6,905	273	58
12	cFBN1_N5256_F cFBN1_N5545_R	CCA GGC TTT GTC ATC GAC AT CAT TGC ACT GTC CTG TGG AG	42 44	7,435	289	64
13	cFBN1_N5817_F cFBN1_N6067_R	AAA TGG GAA TCT TTG CAG AAA TG TTT CTG GCT CTT CGA CAC ACT C	47 49	3,409	257	58
14	cFBN1_N6213_F cFBN1_N6579_R	TCA CCC AAA TCC AGA AAT CAC TC CTC GCA GGT GCA TTC AAA AC	50 53	2,871	366	58
15	cFBN1_N7077_F cFBN1_N7349_R	AAA TCG GAA TGC TGC TGT GA TGG TTG CAC TCG TTC AGA TCT AC	57 58/59	1,702	272	58
16	cFBN1_N7141_F cFBN1_N7522_R	AGG GGA CTG TGG CTT TCA AGA TGA AGC CGC CAA TGG TGT T	57 60	5,249	381	62
17	cFBN1_N7732_F cFBN1_N8075_R	AGC ATG GCT GCC AGA ACA TGC CCA TTC CAG AAA CAC AG	62 64	7,711	342	58

Contributions of authors to the manuscript "Quantitative Sequence Analysis of *FBN1* Premature Termination Stop Codons Provide Evidence for Incomplete NMD in Leukocytes"

István Magyar	Planning and performing experiments, preparation of RNA, primer design, RT-PCR, sequencing, interpretation of data, writing of the manuscript
Eliane Arnold	Mutation analyses
Dvora Colman	Preparation of RNA, RT-PCR, sequencing
Daniela Baumgartner	Sending of patients' material and clinical data, editing of the manuscript
Armand Bottani	Sending of patients' material and clinical data, editing of the manuscript
Siv Fokstuen	Sending of patients' material and clinical data, editing of the manuscript
Marie-Claude Addor	Sending of patients' material and clinical data, editing of the manuscript
Wolfgang Berger	Conceptual input, interpretation of data, editing of the manuscript
Beat Steinmann	Sending of patients' material and clinical data, editing of the manuscript
Thierry Carrel	Sending of patients' material and clinical data, editing of the manuscript
Gábor Mátyás	Conceptual planning and input, design and supervision of the study, interpretation of data, writing and editing of the manuscript

3.3 Large Genomic Fibrillin-1 (*FBN1*) Gene Deletions Provide Evidence for True Haploinsufficiency in Marfan Syndrome

Manuscript published in Human Genetics, August 2007, 122(1):23-32

Gábor Mátyás[1†], Sira Alonso[1†], Andrea Patrignani[2], Myriam Marti[2], Eliane Arnold[3], István Magyar[1], Caroline Henggeler[1], Thierry Carrel[4], Beat Steinmann,[3] and Wolfgang Berger[1]

[1] Division of Medical Molecular Genetics and Gene Diagnostics, Institute of Medical Genetics, University of Zurich, Zurich, Switzerland

[2] Functional Genomics Center Zurich, ETH and University of Zurich, Zurich, Switzerland

[3] Division of Metabolism and Molecular Pediatrics, University Children's Hospital, Zurich, Switzerland

[4] Clinic for Cardiovascular Surgery, University Hospital, Berne, Switzerland

Correspondence to:

Gábor Mátyás, Ph.D., Division of Medical Molecular Genetics and Gene Diagnostics, Institute of Medical Genetics Institute of Medical Molecular Genetics, University of Zurich, Schorenstrasse 16, CH-8603 Schwerzenbach, Switzerland; e-mail: matyas@medgen.uzh.ch

[†]Gábor Mátyás and Sira Alonso have contributed equally to this work.

Hum Genet (2007) 122:23–32
DOI 10.1007/s00439-007-0371-x

ORIGINAL INVESTIGATION

Large genomic fibrillin-1 (*FBN1*) gene deletions provide evidence for true haploinsufficiency in Marfan syndrome

Gábor Mátyás · Sira Alonso · Andrea Patrignani · Myriam Marti · Eliane Arnold · István Magyar · Caroline Henggeler · Thierry Carrel · Beat Steinmann · Wolfgang Berger

Received: 15 March 2007 / Accepted: 18 April 2007 / Published online: 10 May 2007

Abstract Mutations in the *FBN1* gene are the major cause of Marfan syndrome (MFS), an autosomal dominant connective tissue disorder, which displays variable manifestations in the cardiovascular, ocular, and skeletal systems. Current molecular genetic testing of *FBN1* may miss mutations in the promoter region or in other noncoding sequences as well as partial or complete gene deletions and duplications. In this study, we tested for copy number variations by successively applying multiplex ligation-dependent probe amplification (MLPA) and the Affymetrix Human Mapping 500 K Array Set, which contains probes for ~500,000 single-nucleotide polymorphisms (SNPs) across the genome. By analyzing genomic DNA of 101 unrelated individuals with MFS or related phenotypes in whom standard genetic testing detected no mutation, we identified *FBN1* deletions in two patients with MFS. Our high-resolution approach narrowed down the deletion breakpoints. Subsequent sequencing of the junctional fragments revealed the deletion sizes of 26,887 and 302,580 bp, respectively. Surprisingly, both deletions affect the putative regulatory and promoter region of the *FBN1* gene, strongly indicating that they abolish transcription of the deleted allele. This expectation of complete loss of function of one allele, i.e. true haploinsufficiency, was confirmed by transcript analyses. Our findings not only emphasize the importance of screening for large genomic rearrangements in comprehensive genetic testing of *FBN1* but, importantly, also extend the molecular etiology of MFS by providing hitherto unreported evidence that true haploinsufficiency is sufficient to cause MFS.

Gábor Mátyás and Sira Alonso have contributed equally to this work.

G. Mátyás (✉) · S. Alonso · I. Magyar · C. Henggeler · W. Berger
Division of Medical Molecular Genetics and Gene Diagnostics,
Institute of Medical Genetics, University of Zurich,
Schorenstrasse 16, 8603 Schwerzenbach, Zurich, Switzerland
e-mail: matyas@medgen.unizh.ch

A. Patrignani · M. Marti
Functional Genomics Center Zurich,
ETH and University of Zurich, Zurich, Switzerland

E. Arnold · B. Steinmann
Division of Metabolism and Molecular Pediatrics,
University Children's Hospital, Zurich, Switzerland

T. Carrel
Clinic for Cardiovascular Surgery,
University Hospital, Berne, Switzerland

Introduction

Marfan syndrome (MFS; MIM# 154700) is a prevalent connective tissue disorder with variable manifestations in the skeletal, ocular, and cardiovascular systems (De Paepe et al. 1996). MFS is inherited in an autosomal dominant manner and caused by mutations in the gene encoding the 350-kD extracellular matrix protein fibrillin-1 (*FBN1*; MIM# 134797) in the majority of cases (Dietz et al. 1991). Recently, heterozygous mutations in the genes coding for transforming growth factor beta receptors I (*TGFBR1*; MIM# 190181) and II (*TGFBR2*; MIM# 190182) have also been reported in patients with MFS-related disorders, such as MFS type 2 (MFS2; MIM# 154705), Loeys-Dietz aortic aneurysm syndrome (LDS; MIM# 609192), and familial thoracic aortic aneurysms and dissections (TAAD; MIM# 132900), indicating genetic heterogeneity in MFS and its related conditions (Mizuguchi et al. 2004; Loeys et al. 2005; Pannu et al. 2005; Mátyás et al. 2006; Loeys et al. 2006).

In classical MFS patients, mutation analyses have failed to detect *FBN1* involvement in at least 10% of cases,

suggesting not only that the disease-causing mutation occurs in a different gene but also demonstrate the limitations of commonly used standard PCR-based screening approaches. Indeed, current molecular genetic testing of *FBN1*, although powerful, may miss mutations in the promoter region or in other noncoding sequences as well as deletions/duplications of the entire gene or part of it. Accordingly, in the locus-specific mutation databases more than 600 unique *FBN1* mutations but only 12 (2.0%) large deletions have been registered (UMD, http://www.umd.be:2030; HGMD, http://www.hgmd.cf.ac.uk/ac/gene.php?gene = FBN1). In contrast, the relative frequency of large deletions available in the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk) containing 64,251 mutations in 2,362 genes (version 09/12/2006) is significantly higher (5.6%), suggesting that more large *FBN1* deletions await detection.

In this study, we tested this hypothesis by applying the recently introduced multiplex ligation-dependent probe amplification (MLPA) technique, which is suitable for the identification of large deletions and less laborious than methods traditionally used for detection of gross rearrangements, such as Southern blot and cytogenetic techniques. MLPA relies on sequence-specific probe hybridization to genomic DNA, followed by amplification of the hybridized probe with universal primers, and semi-quantitative analysis of the resulting PCR products (Schouten et al. 2002). Here, we report on large deletions affecting the 5′ region of the *FBN1* gene in two unrelated MFS patients. The breakpoints of the deletions identified by MLPA were narrowed down by means of high-density single-nucleotide polymorphism (SNP) arrays and characterized by using long-range PCR, subsequent cycle sequencing, and transcript analyses. Our data provide hitherto unreported evidence for true haploinsufficiency in patients with MFS.

Subjects and methods

Patients

A cohort of 101 unrelated patients with suspected MFS (~40%) or MFS-like phenotypes (~60%) involving monosymptomatically or predominantly the cardiovascular (~39%), skeletal (~18%) or ocular (~3%) system was selected for this study. In this cohort, previous denaturing high-performance liquid chromatography (DHPLC) analysis and/or sequencing of all exons and flanking intronic regions of *FBN1*, *TGFBR1*, and *TGFBR2* revealed no disease-causing sequence variant (Mátyás et al. 2002a, 2006). Data on the clinical phenotypes of patients were collected from medical records or during physical examinations by one of the authors (B.S, T.C.).

Multiplex ligation-dependent probe amplification (MLPA)

A total of 100 ng of genomic DNA of each patient was screened by MLPA using the SALSA kits P065 (probes for *FBN1* and *TGFBR2*), P066 (probes for *FBN1*), and P148 (probes for both *TGFBR1* and *TGFBR2*), commercially available from MRC-Holland (Amsterdam, The Netherlands), on an ABI PRISM 310 or 3100 Genetic Analyzer (Applied Biosystems, Rotkreuz, Switzerland) according to the manufacturers' instructions. Each MLPA signal was normalized and compared to the corresponding peak area obtained in a control DNA sample. Deviations more than 30% were suspected as alterations and verified by repeated MLPA analysis.

High-density microarray analyses

In order to narrow down the breakpoints of the deletions identified by MLPA, we used the high-density GeneChip Human Mapping 500 K Array Set (Affymetrix, Santa Clara, CA, USA), which contains ~500,000 genomewide SNPs, according to the manufacturer's manual. After hybridization and scanning, raw data were processed using the GCOS 1.4 software (Affymetrix, Santa Clara, CA, USA) and probe cell intensities were calculated and summarized for the respective probe sets by means of the MAS5 algorithm (Hubbell et al. 2002). SNP allele calls, marker intensity values, and quality control measures were computed by the GTYPE software (Affymetrix, Santa Clara, CA, USA) using the integrated RLMM algorithm (Rabbee and Speed 2006). Quality control measures were considered before performing the statistical analysis, including appropriate numbers of present calls (>97%) and control SNP performance (Modified Partitioning Around Medoids [MPAM] algorithm Call Rate [MCR] close to or higher than 90%). For the determination of regions with loss of heterozygosity and decreased signal intensity, advanced data analyses were performed using the academic software tools dChip (Harvard University; http://www.biosun1.harvard.edu/complab/dchip) and CNAG2.0 (University of Tokyo; http://www.genome.umin.jp).

Amplification and sequencing of the junctional fragments

Primers flanking the predicted deletions were designed based on decreased MLPA and array signal intensities and were used in long-range PCR (44F 5′-ACGAACCTTTCAAATATTCCCCC-3′, 44R 5′-TGGGAGTGATGGGACAACTGAG-3′, 70F 5′-ACTGCTGCAGAGTGCCTGATA-3′, 70R 5′-GGGCTATGTGCTCTGGCTATT-3′). Briefly, 100 ng of DNA and Buffer 3 were used with the Expand Long Template PCR System (Roche Diagnostics, Rotkreuz, Switzerland) according to the manufacturer's instructions

Hum Genet (2007) 122:23–32

with the following cycling profile: 5 min initial denaturation at 92°C; 10 cycles of 10 s at 92°C, 30 s at 58.5°C (for 44F/44R) or 59.3°C (for 70F/70R), and 8 min at 68°C; followed by 25 cycles of 15 s at 92°C, 30 s at 58.5°C (for 44F/44R) or 59.3°C (for 70F/70R), and 8 min (+20 s/cycle) at 68°C; finished by a 7 min final extension step at 68°C. Amplicons were treated with ExoSAP-IT (USB Corporation, Cleveland, OH, USA) and sequenced using internal primers by means of a BigDye Terminator Cycle Sequencing Ready Reaction kit v1.1 on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Rotkreuz, Switzerland). Sequences were analyzed using SeqScape 2.5 (Applied Biosystems, Rotkreuz, Switzerland) and the genome browser of the University of California, Santa Cruz (UCSC; http://www.genome.ucsc.edu). Sequence variants were referenced with respect to the human genome reference sequence (NCBI build 36.1, March 2006) according to the guidelines of the HGVS (http://www.hgvs.org/mutnomen).

Transcript analyses

Primary fibroblasts were available only from Patient 70. Total RNA was isolated by the RNeasy mini kit (Qiagen, Hilden, Germany) from cultured fibroblasts. Prior to RNA extraction, fibroblasts were incubated with 10 µg/ml cycloheximide (Actidione; Sigma, St. Louis, MO, USA) for 15 h in order to inhibit potential nonsense-mediated mRNA decay. Reverse transcription (RT)-PCR was used to amplify a complementary DNA (cDNA) fragment (1,484 bp) containing the polymorphism c.8,930C > T (rs1042078, average heterozygosity 0.490 ± 0.071, dbSNP build 126) that is located in the 3′ untranslated region (UTR) of the *FBN1* gene and was detected during the mutation screening at genomic DNA level. RT-PCR was performed by means of the OneStep RT-PCR kit (Qiagen, Hilden, Germany) using 125 ng RNA and primers specific to the *FBN1* exons 63 (5′-AGGGCGGTTACCTGTGTG-3′) and 65 (5′-CCCCT TGTTGACAGGAATGAC-3′). RT-PCR product was sequenced by using an internal primer in exon 65 (5′-AGC ACCATTACAAACCCTCACA-3′) as described above. The presence of allele-specific transcripts was considered according to the procedure previously outlined (Mátyás et al. 2002b; Qiu et al. 2003).

Results

Multiplex ligation-dependent probe amplification (MLPA)

In 2 out of 101 previously "mutation-negative" patients with MFS or related phenotypes, MLPA analyses resulted in abnormal patterns. In Patient 44, MLPA revealed reduced relative peak areas for both probes of *FBN1* exon 1 (Fig. 1), suggesting the deletion of this exon. Only the father's DNA was available for MLPA analysis, which revealed a normal profile (data not shown). In Patient 70, MLPA showed reduced relative peak areas of fragments corresponding to exons 1–9 and 12–16 (Fig. 1), suggesting the deletion of exons 1–16. The parents of Patient 70 were not available for testing. All other patients showed relative peak areas within the range defined as normal. MLPA for the *TGFBR1* and *TGFBR2* genes (kit P148) showed normal relative peak areas in all patients (data not shown). Taken together, the relative frequency of gross *FBN1* deletions in patients with suspected MFS in whom exon-by-exon screening revealed no mutation in the genes *FBN1*, *TGFBR1*, and *TGFBR2* can be estimated as 2/101 (0.3–7.7%, $P = 0.05$; according to VassarStats, http://www.faculty.vassar.edu/lowry/prop1.html). Patients 44 and 70 were sporadic cases and fulfilled the diagnostic criteria of Ghent nosology (Table 1).

Identification and characterization of breakpoints

Loss of heterozygosity and decreased signal intensities upon high-density SNP array analyses confirmed the MLPA results and narrowed down the deletion breakpoints. In Patient 44, loss of heterozygosity occurred in a region spanning 94.5 kb between the SNPs rs1678982 and rs784408, while reduction of signal intensities indicated a deletion of less than 44.4 kb between rs2018854 and rs931781 (Fig. 2a). Similarly, in Patient 70 the deletion was localized to a region of ~317 kb between rs11070644 and rs2899422 by considering decreased array signal intensities (Fig. 3a).

Based on these data, forward and reverse primers were designed for long-range PCR, which resulted in fragments of ~6.5 and ~8.5 kb in Patient 44 and Patient 70, respectively. Due to the size of normal alleles (33.2 kb using primers 44F/44R and 311.3 kb with primers 70F/70R), amplifications were observed from the deleted alleles only, but not from the normal alleles of either Patient 44, his father, Patient 70, or several controls (data not shown). Subsequent sequencing of the long-range PCR products identified a deletion of 26,887 bp in Patient 44 and 302,580 bp in Patient 70 (Figs. 2b, 3b).

Transcript analyses

Both deletions affect exon 1 (345 bp), which contains the translation initiation ATG codon at nucleotide positions 182–184, and a ~4-kb upstream region of the *FBN1* gene, which may harbor the promoter, thus strongly indicating that they abolish the transcription/translation of the respective deleted allele (Figs. 2, 3). This expectation of complete loss of function of one allele (true haploinsufficiency) was confirmed by RT-PCR sequence analysis in Patient 70 being

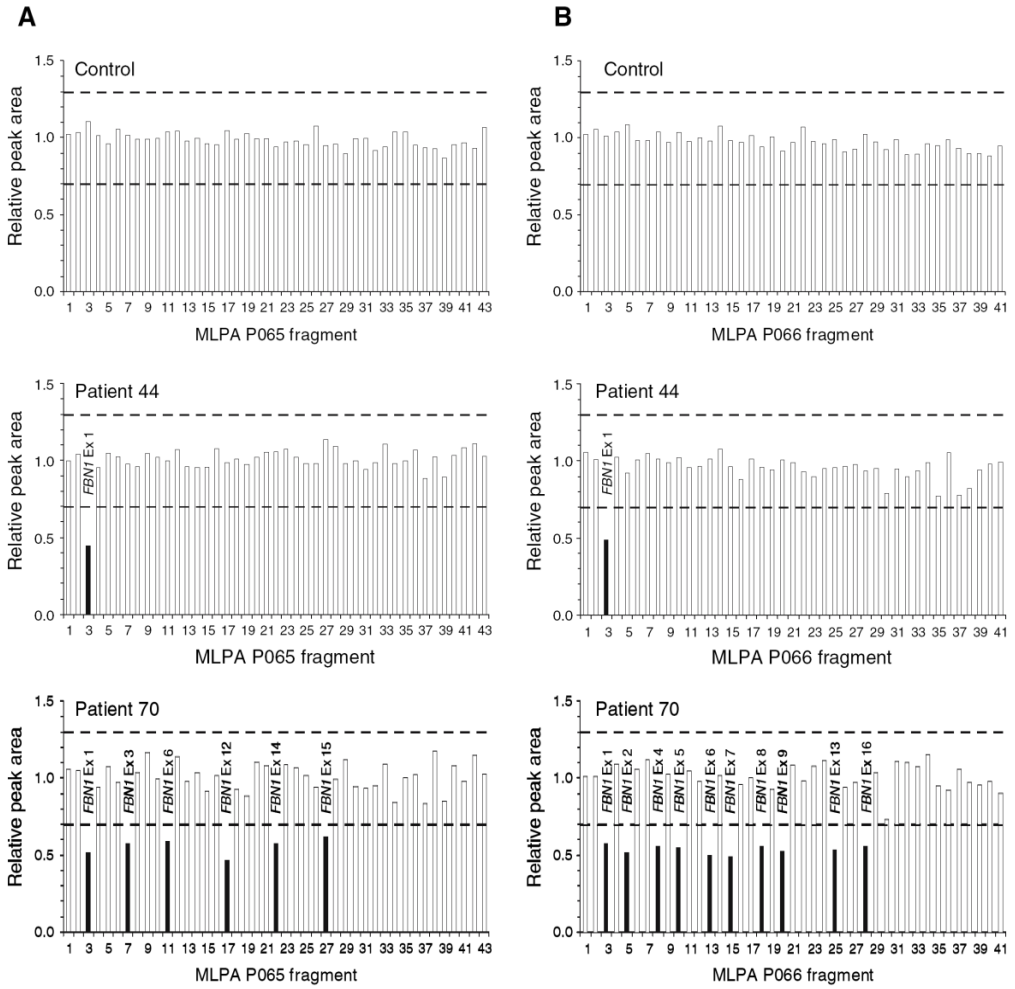

Fig. 1 Results of semiquantitative MLPA analyses in Patients 44 and 70. The normal range of normalized relative peak areas (*white bars*) is given by *dotted lines* (±0.30). Values out of this range are marked by *black bars* and numbers of the corresponding *FBN1* exons (Ex). **a** Normalized relative peak areas measured with the P065 kit, which consists of 25 MLPA probes for *FBN1* and eleven control MLPA probes located on chromosomes 1, 2, 4, 5, 8, 11, and 16 (fragments 1, 2, 6, 13, 19, 25, 31, 32, 36, 40, and 43) as well as probes for all seven *TGFBR2* exons (fragments 8, 12, 16, 20, 24, 28, and 34). **b** Normalized relative peak areas measured with the P066 kit, which consists of MLPA probes for 30 *FBN1* exons, in two of which (exons 1 and 6) a probe of P065 is located as well, and eleven control MLPA probes located on chromosomes 1, 3, 5, 6, 13, 16, and 22 (fragments 1, 2, 7, 12, 17, 22, 27, 31, 36, 40, and 41). Taken together the results of the P065 and P066 kits, MLPA revealed reduced relative peak areas of fragments corresponding to the *FBN1* exon 1 in Patient 44 and to the *FBN1* exons 1–9 and 12–16 in Patient 70 (note that there are no MLPA probes for the *FBN1* exons 10 and 11 as well as 20, 22, 27 32, 37, 39, 48, 51, 59, and 61)

heterozygous for the c.8,930C > T polymorphism in the *FBN1* 3′UTR, which showed the transcription of only one allele (Fig. 4). Appropriate clinical samples were not available from Patient 44 to verify this finding in affected tissue.

Discussion

In this study, we applied for the first time MLPA and high-density SNP arrays to analyze genomic DNA samples of patients with MFS or related phenotypes in whom standard molecular testing detected no mutation. Our results contribute to the molecular etiology of MFS by providing previously unprecedented evidence for true haploinsufficiency in patients. Our data significantly extend the number of large *FBN1* deletions, which have previously been detected by cDNA or Southern blot analyses (Table 2), demonstrating that comprehensive genetic testing of *FBN1* should include screening for large genomic rearrangements by using an appropriate method.

The P065-P066 MLPA kit used in this study contains probes for 53 of the 65 *FBN1* exons. Since not all *FBN1*

Hum Genet (2007) 122:23–32

Table 1 Clinical profiles of patients carrying a large deletion identified in this study

Organ systems	Criteria[a]	Patient 44	Patient 70
Skeletal	*Major criteria*		
	Pectus carinatum	−	+
	Reduced upper-segment to lower-segment ratio or arm span to height ratio >1.05	+	+
	Wrist and thumb signs	+	+
	Scoliosis of >20°	+	−
	Reduced extension at the elbows (<170°)	+	+
	Minor criteria		
	Pectus excavatum of moderate severity (asymmetric)	+	−
	Joint hypermobility	+	−
	Highly arched palate with crowding of teeth	+	+
	Facial appearance (retrognathia, malar hypoplasia, down-slanting palpebral fissures)	+	+
	Nonfamilial overgrowth	+	+
Ocular	*Major criterion*		
	Ectopia lentis	−	−
	Minor criterion		
	High myopia	+	+
Cardiovascular	*Major criterion*		
	Dilatation of the ascending aorta	+	+
	Minor criterion		
	Mitral valve prolapse with mitral valve regurgitation	+	+
Skin and integument	*Minor criterion*		
	Striae distensae without obvious cause	+	−
Pulmonary		NI	NI
Dura		NP	NP
Family history		NI	NI

+ Present; − absent; *NP* not performed; *NI* system not involved

[a] Only criteria present in one of the patients are listed (except major criterion in the ocular system)

exons were analyzed, a deletion or duplication of non-tested exons cannot be excluded. Furthermore, copy-number neutral rearrangements as well as those affecting no MLPA-specific sequences could not be detected because such mutations do not change the relative peak area of MLPA probes. In contrast to traditionally used techniques, MLPA is a simple method to detect gross deletions. However, sequence variations in the probe-binding regions as well as probe and DNA quality may significantly affect MLPA results. Accordingly, it is highly recommended to confirm MLPA results by an independent method. Both deletions described here were also detectable by quantitative real-time PCR (data not shown). Like MLPA, the Mapping 500 K Array Set cannot detect balanced rearrangements and is sensitive to handling and DNA quality in our experience. Deletions that affect only a small proportion of templates (mosaicism) would also be expected to present detection problems for both MLPA and SNP array technologies.

The deletions presented here generated junction fragments with short stretches of identical sequences at the sites of breakpoints (Figs. 2b, 3b), a phenomenon that has been described not only in *FBN1* but also in other genes (Giacalone and Francke 1992; Otto et al. 2000; Liu et al. 2001). The precise mechanism of this phenomenon, however, remains to be elucidated. Both deletions comprise exon 1 and a ~4-kb upstream region of *FBN1*, a surprising finding not reported before, which is most likely due to the limitations of previously used screening methods. A priori, one might expect that these deletions affect the regulatory or promoter regions of *FBN1*. Indeed, both deleted regions include the three alternatively spliced exons of the *FBN1* gene (C, A, and B; Fig. 2a), which are thought to have regulatory function (Corson et al. 1993; Biery et al. 1999). Furthermore, although little is known about the *FBN1* promoter region, very recent studies indicates that the *FBN1* gene is controlled by a single promoter with a start site located up to 500 bases upstream (5′) of the ATG start codon (11th International Congress of Human Genetics, http://www.ichg2006.com/abstract/488.htm), which is affected by both deletions as well.

In general, no transcript will be produced when the promoter is deleted, unless alternative down- or up-stream

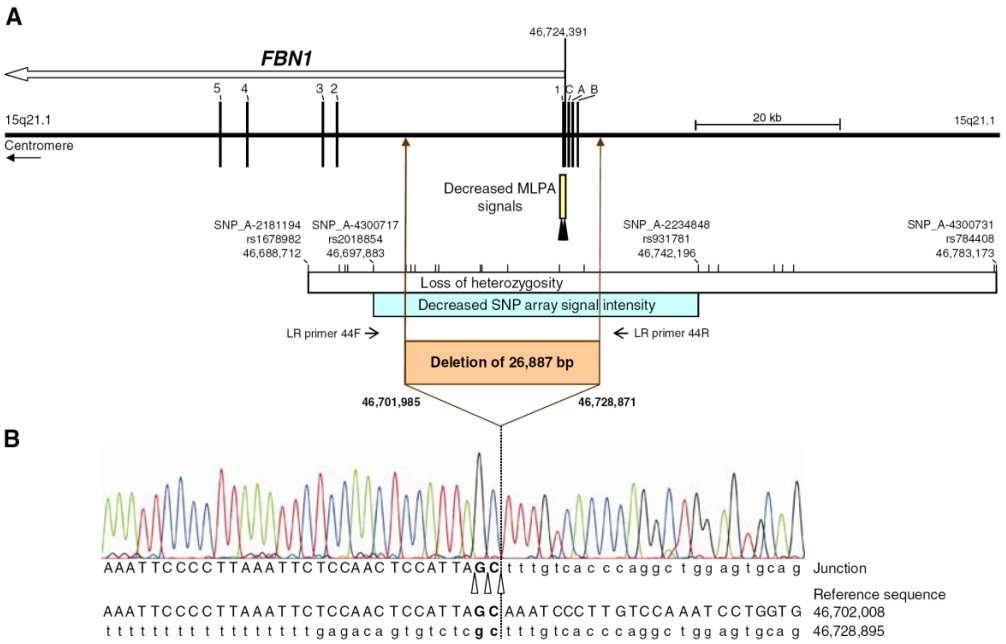

Fig. 2 Breakpoint analyses in Patient 44. **a** Schematic representation of the genome in the region of deletion as well as an overview of the results of the MLPA and SNP array analyses. The *open arrow* below the gene name indicates the direction of transcription. Exons are specified by bars and marked with the corresponding number or character (cf. *C*, *A*, and *B* are alternatively spliced exons of *FBN1*, while exon 1, also known as exon M, harbors the presumptive translation initiation codon). The region with decreased normalized MLPA signals (Fig. 1) is indicated by a *yellow bar* and the positions of the two MLPA probes located in exon 1 of *FBN1* are denoted by *arrows* (*filled triangle*). The regions with loss of heterozygosity and decreased SNP array signal intensity both gained by the Affymetrix GeneChip Human Mapping 500 K Array Set are indicated by *white* and *blue bars*, respectively (note that the flanking SNPs are not included in the regions, displaying an open interval). The positions of SNPs tested in the array set are denoted by *vertical lines*. Note the large region (~67 kb) with loss of heterozygosity not associated with the deletion. Primers designed for long-range PCR (LR primers 44F and 44R) and the deleted region of the genomic DNA (*brown bar*) are indicated. **b** Sequences of the long-range PCR product spanning the breakpoint junction of the deletion. *Uppercase letters* indicate sequences of *FBN1* intron 1, *lowercase letters* denote intergenic sequences. Due to identical sequences at the site of breakpoints, the break and rejoining could have occurred at three positions as indicated by *arrows* (*open triangle*). The *dotted line* marks the most telomeric position of the possible breakpoints. All nucleotide positions are described in relation to the human genome reference sequence (NCBI build 36.1)

promoter is activated. The latter occurs when a deletion fuses two genes that are in the same transcriptional orientation. Since the deletion in Patient 70 involves also the neighboring *CEP152* gene, which encodes the centrosomal protein of 152 kDa with unknown function (Fig. 3a), we made use of a polymorphism in the *FBN1* 3′UTR to determine whether or not both alleles are transcribed. In Patient 70, sequence analysis of RT-PCR products revealed transcripts from only one allele (Fig. 4). This finding is consistent with the expectation that the deletion involves the regulatory and promoter regions and thus prevents transcription. The impact of the partial deletion of *CEP152*, if any, is unknown. In Patient 44, the activation of an alternative cryptic promoter, e.g. downstream of the deletion in intron 1 of *FBN1*, could not be excluded. The quantification of the total amount of *FBN1* transcripts was not within the scope of this study, but, similar to other modifying factors, differences in normal *FBN1* expression may explain some of the clinical variability seen in Patients 44 and 70 (Table 1), as previously demonstrated in a family with MFS (Hutchinson et al. 2003). Although, except of the father of Patient 44, neither the parents nor other family members (with no hints for MFS by history) were available for genetic testing, negative family histories suggest de novo occurrence of the deletions.

Interestingly, neither Patient 44 nor Patient 70 presents any sign of ectopia lentis. This is in accordance with the observation that MFS patients with a premature termination codon (PTC) mutation, which can lead to functional haploinsufficiency, i.e., preferential degradation of mutant transcripts due to nonsense-mediated mRNA decay (NMD), have a distinctly lesser chance of developing ectopia lentis than individuals with *FBN1* cysteine substitutions (Schrijver et al. 2002; Rommel et al. 2005). However, the clinical findings in Patients 44 and 70 could also be explained by the lack of ocular involvement seen in about one third of

Hum Genet (2007) 122:23–32

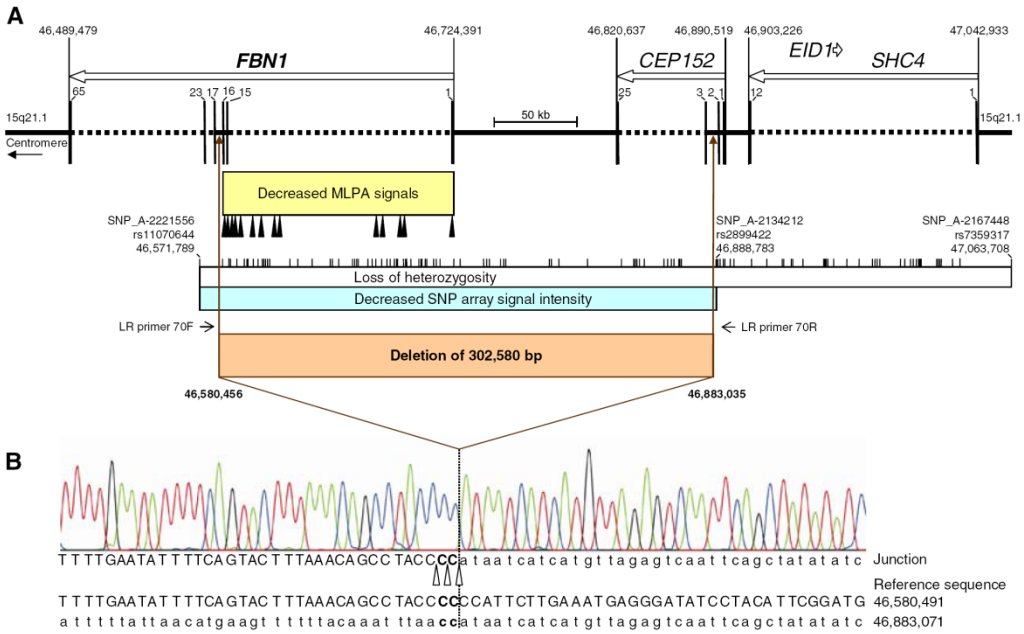

Fig. 3 Breakpoint analyses in Patient 70. Symbols and labels are as used in Fig. 2. **a** Schematic representation of the genome in the region of deletion as well as overview of the results of the MLPA and SNP array analyses. Note the extended region (~180 kb) with loss of heterozygosity not deleted. Forward and reverse primers designed for long-range PCR are indicated (LR primers 70F and 70R). For clarity, exons C, A, and B of *FBN1* are not shown (cf. Fig. 2). **b** Sequences of the long-range PCR product spanning the breakpoint junction of the deletion. *Uppercase letters* indicate sequences of *FBN1* intron 16, and *lowercase letters* denote sequences of *CEP152* intron 2

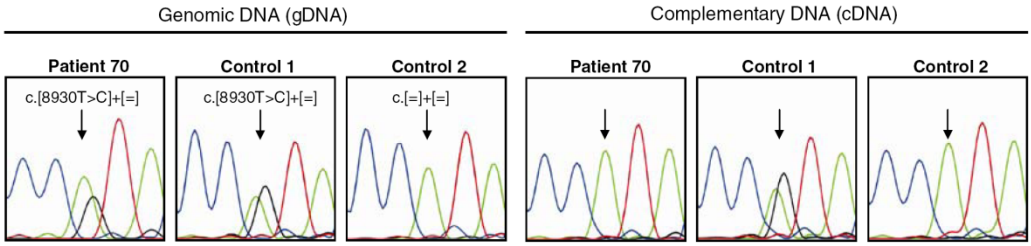

Fig. 4 Partial reverse sequence of the *FBN1* 3′ untranslated region flanking the polymorphism c.8,930C > T (g.46,490,165 NCBI build 36.1; *arrowed*). Note that the cDNA sequence of Patient 70 heterozygous for c.8,930C > T lacks allele C (*black*) and is indistinguishable from the cDNA sequence of Control 2 homozygous for allele T (*green*), while in the controls the results at gDNA and cDNA levels are consistent with respect to each other. Complementary DNA sequences were derived from cultured fibroblast mRNA stabilized by the translation inhibitor cycloheximide

MFS patients. In this context, mutations resulting in functional haploinsufficiency can also be associated with mild phenotypes which may fail to meet the diagnostic criteria for MFS (Dietz et al. 1993). However, such association between mutant transcript abundance and disease severity could not be found in Patients 44 and 70, similar to previously reported patients with classic severe MFS who harbored very low levels of PTC-containing mutant transcripts (Halliday et al. 1999).

Taken together, Patient 70, and most likely Patient 44 as well, have only a single functional *FBN1* allele, representing true haploinsufficiency. Since the first description of an *FBN1* mutation (Dietz et al. 1991), the role of haploinsufficiency in the pathogenesis of MFS has been the subject of intensive investigations and discussions. Mouse models suggested a critical threshold of functional microfibrils in the disease presentation of MFS (Pereira et al. 1999) and provided evidence for a critical contribution of haploinsufficiency by showing that half-normal amounts of fibrillin-1 can be insufficient to initiate productive microfibrillar assembly (Dietz and Mecham 2000; Judge et al. 2004). Furthermore, mouse MFS models showed that mice

Table 2 Overview of cases with large genomic deletions affecting the *FBN1* gene

Deletion (*FBN1* exon affected)[a]	Detection rate[b]	Patient		Reference (Year)
		Age (years)	Phenotype[c]	
g.46,701,985_46,728,871del26,887 (1)[d]	1/101	25	See Table 1	Patient 44 in this study
g.46,580,456_46,883,035del302,580 (1–16)[d]	1/101	40	See Table 1	Patient 70 in this study
FBN1:g.182,966_189,399del6,434 (42–43)[e]	1/60	>45	Classic MFS	Liu et al. (2001)
FBN1:g.192,760_199,897del7,138 (44–46)[f]	1/60	0	Neonatal MFS	Liu et al. (2001)
FBN1:p.Asp2127_Val2165del (52)[g]	1/93	40	Classic MFS	Loeys et al. (2004)
FBN1:c.7,456_7,821del366 (60–62)[h]	1/61	48	Classic MFS	Kainulainen et al. (1992)
FBN1:g.219,771_231,247del11,477 (58–63)[i]	1/11	17	Classic MFS	Singh et al. (2007)

[a] Only deletions not detectable by standard exon-by-exon screening methods on genomic DNA level are shown. Cytogenetically detectable deletions encompassing the *FBN1* gene, such as deletions involving 15q21.1, 15q or the entire chromosome 15 (Fukushima et al. 1990; Dierlamm et al. 2003; Shur et al. 2003; Pramparo et al. 2005), are not included here because these chromosomal abnormalities affect not only *FBN1* but also several other genes. For similar reason, entire *FBN1* allele deletions reported by Hutchinson et al. (2003) and Adès et al. (2006) as well as mentioned by Judge et al. (2004) are not included. In accordance with the guidelines of the HGVS (http://www.hgvs.org/mutnomen), in the description of deletions affecting only the *FBN1* gene (FBN1:) position +1 corresponds to the first nucleotide of the *FBN1* reference sequence (GenBank NC_000015.8) at genomic DNA (g) level, to the A of the ATG start codon of the mRNA reference sequence (GenBank NM_000138.3) at cDNA (c) level and to the translation initiator methionine of the reference sequence (GenBank NP_000129.2) at protein (p) level, respectively

[b] Detection rate (#1/#2) is given as the number of cases with the respective mutation (#1) among patients screened in the corresponding study (#2)

[c] *MFS* Marfan syndrome

[d] Described in relation to the human genome reference sequence (NCBI build 36.1, March 2006), starting with the first-deleted and ending with the last-deleted nucleotide. See also Figs. 2 and 3

[e] The nucleotide at position +325 in intron 41 is joined to the nucleotide in position +314 in intron 43 (Liu et al. 2001)

[f] The breakage and rejoining occurred within the pentamer (atttt) at positions −282 to −278 of intron 43 and positions −325 to −321 of intron 46 (Liu et al. 2001)

[g] In the study of Loeys et al. (2004), this deletion involving exon 52 was not identified by mutation screening of *FBN1* genomic DNA but by cDNA analyses and was described solely at the protein level as presented here

[h] This deletion was originally published by Kainulainen et al. (1992) as c.4,762_5,127del366 using an at that time incomplete *FBN1* cDNA sequence for the description

[i] The nucleotide at position −88 in intron 57 is joined to the nucleotide in position +1,062 in intron 63 (Singh et al. 2007)

haploinsufficient in fibrillin-1 have marked dysregulation of transforming growth factor beta (TGF-beta) signaling (Neptune et al. 2003), most likely due to the fact that fibrillin-1 regulates the bioavailability of TGF-beta 1 (Chaudhry et al. 2007). In humans, although several observations supported a traditional dominant-negative mechanism, it has been suggested that functional haploinsufficiency through NMD of most of the mutant mRNAs plays a role in the pathogenic mechanism of MFS (Dietz et al. 1993; Hewett et al. 1994; Nijbroek et al. 1995; Pepe et al. 2001; Caputi et al. 2002; Judge and Dietz 2005; Robinson et al. 2006). However, since mutant transcripts being subject to NMD are often not completely degraded, it has remained unclear, whether the potent dominant-negative effect of the remaining mutant transcripts leads to the disease. Even cytogenetically detectable heterozygous deletions of the entire *FBN1* gene, although suggested, failed to demonstrate the role of true haploinsufficiency in MFS, because these chromosomal abnormalities affect not only *FBN1* but also several other genes, resulting in different phenotypes do not meet the Ghent criteria (e.g., Hutchinson et al. 2003; Adès et al. 2006). Hence, the findings in this study demonstrate for the first time that true haploinsufficiency, i.e., the complete loss of function of one *FBN1* allele, is sufficient to cause MFS in patients. In such cases, increased expression from the normal *FBN1* allele may represent a novel therapeutic avenue.

Acknowledgments We are grateful to the patients, the father of Patient 44, and the referring physicians for participating in this study. We thank Marzanna Künzli, Ulrich Wagner, Philippe Reuge for technical support and assistance; Angelika Schwarze for cell cultures; Melanie Maudrich for initial mutation analysis of the *FBN1* gene; members of the Institute of Medical Genetics, University of Zurich, for constructive discussions. This work was supported by the Functional Genomics Center Zurich and grants from the Foundation for Research at the Medical Faculty and Research Funding of the University of Zurich (to G.M.), Swiss Heart Foundation (to G.M.), Swiss National Science Foundation (NF 3200B0-109370/1 to B.S.), and the Wolfermann-Nägeli-Stiftung Zurich (to B.S.).

References

Adès LC, Sullivan K, Biggin A, Haan EA, Brett M, Holman KJ, Dixon J, Robertson S, Holmes AD, Rogers J, Bennetts B (2006) *FBN1*, *TGFBR1*, and the Marfan-craniosynostosis/mental retardation disorders revisited. Am J Med Genet A 140:1047–1058

Biery NJ, Eldadah ZA, Moore CS, Stetten G, Spencer F, Dietz HC (1999) Revised genomic organization of *FBN1* and significance for regulated gene expression. Genomics 56:70–77

Caputi M, Kendzior RJ Jr, Beemon KL (2002) A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer. Genes Dev 16:1754–1759

Chaudhry SS, Cain SA, Morgan A, Dallas SL, Shuttleworth CA, Kielty CM (2007) Fibrillin-1 regulates the bioavailability of TGFbeta1. J Cell Biol 176:355–67

Corson GM, Chalberg SC, Dietz HC, Charbonneau NL, Sakai LY (1993) Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure and alternatively spliced exons at the 5′ end. Genomics 17:476–484

De Paepe A, Devereux RB, Dietz HC, Hennekam RCM, Pyeritz RE (1996) Revised diagnostic criteria for the Marfan syndrome. Am J Med Genet 62:417–426

Dierlamm J, Schilling G, Michaux L, Hinz K, Penas EM, Seeger D, Hagemeijer A, Hossfeld DK (2003) Deletion of chromosome 15 represents a rare but recurrent chromosomal abnormality in myelocytic malignancies. Cancer Genet Cytogenet 144:1–5

Dietz HC, Mecham RP (2000) Mouse models of genetic diseases resulting from mutations in elastic fiber proteins. Matrix Biol 19:481–488

Dietz HC, Cutting GR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, Puffenberger EG, Hamosh A, Nanthakumar EJ, Curristin SM, Stetten G, Meyers DA, Francomano CA (1991) Marfan syndrome caused by a recurrent *de novo* missense mutation in the fibrillin gene. Nature 352:337–339

Dietz HC, McIntosh I, Sakai LY, Corson GM, Chalberg SC, Pyeritz RE, Francomano CA (1993) Four novel *FBN1* mutations: significance for mutant transcript level and EGF-like domain calcium binding in the pathogenesis of Marfan syndrome. Genomics 17:468–475

Fukushima Y, Wakui K, Nishida T, Nishimoto H (1990) Craniosynostosis in an infant with an interstitial deletion of 15q [46,XY,del(15)(q15q22.1)]. Am J Med Genet 36:209–213

Giacalone JP, Francke U (1992) Common sequence motifs at the rearrangement sites of a constitutional X/autosome translocation and associated deletion. Am J Hum Genet 50:725–741

Halliday D, Hutchinson S, Kettle S, Firth H, Wordsworth P, Handford PA (1999) Molecular analysis of eight mutations in *FBN1*. Hum Genet 105:587–597

Hewett D, Lynch J, Child A, Firth H, Sykes B (1994) Differential allelic expression of a fibrillin gene (*FBN1*) in patients with Marfan syndrome. Am J Hum Genet 55:447–452

Hubbell E, Liu WM, Mei R (2002) Robust estimators for expression analysis. Bioinformatics 18:1585–1592

Hutchinson S, Furger A, Halliday D, Judge DP, Jefferson A, Dietz HC, Firth H, Handford PA (2003) Allelic variation in normal human *FBN1* expression in a family with Marfan syndrome: a potential modifier of phenotype? Hum Mol Genet 12:2269–2276

Judge DP, Dietz HC (2005) Marfan's syndrome. Lancet 366:1965–1976

Judge DP, Biery NJ, Keene DR, Geubtner J, Myers L, Huso DL, Sakai LY, Dietz HC (2004) Evidence for a critical contribution of haploinsufficiency in the complex pathogenesis of Marfan syndrome. J Clin Invest 114:172–181

Kainulainen K, Sakai LY, Child A, Pope FM, Puhakka L, Ryhanen L, Palotie A, Kaitila I, Peltonen L (1992) Two mutations in Marfan syndrome resulting in truncated fibrillin polypeptides. Proc Natl Acad Sci USA 89:5917–5921

Liu W, Schrijver I, Brenn T, Furthmayr H, Francke U (2001) Multiexon deletions of the *FBN1* gene in Marfan syndrome. BMC Med Genet 2:11

Loeys B, De Backer J, Van Acker P, Wettinck K, Pals G, Nuytinck L, Coucke P, De Paepe A (2004) Comprehensive molecular screening of the *FBN1* gene favors locus homogeneity of classical Marfan syndrome. Hum Mutat 24:140–146

Loeys BL, Chen J, Neptune ER, Judge DP, Podowski M, Holm T, Meyers J, Leitch CC, Katsanis N, Sharifi N, Xu FL, Myers LA, Spevak PJ, Cameron DE, De Backer J, Hellemans J, Chen Y, Davis EC, Webb CL, Kress W, Coucke P, Rifkin DB, De Paepe AM, Dietz HC (2005) A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in *TGFBR1* or *TGFBR2*. Nat Genet 37:275–281

Loeys BL, Schwarze U, Holm T, Callewaert BL, Thomas GH, Pannu H, De Backer JF, Oswald GL, Symoens S, Manouvrier S, Roberts AE, Faravelli F, Greco MA, Pyeritz RE, Milewicz DM, Coucke PJ, Cameron DE, Braverman AC, Byers PH, De Paepe AM, Dietz HC (2006) Aneurysm syndromes caused by mutations in the TGF-beta receptor. N Engl J Med 355:788–798

Mátyás G, De Paepe A, Halliday D, Boileau C, Pals G, Steinmann B (2002a) Evaluation and application of denaturing HPLC for mutation detection in Marfan syndrome: identification of 20 novel mutations and two novel polymorphisms in the *FBN1* gene. Hum Mutat 19:443–456

Mátyás G, Giunta C, Steinmann B, Hossle JP, Hellwig R (2002b) Quantification of single nucleotide polymorphisms: a novel method that combines primer extension assay and capillary electrophoresis. Hum Mutat 19:58–68

Mátyás G, Arnold E, Carrel T, Baumgartner D, Boileau C, Berger W, Steinmann B (2006) Identification and in silico analyses of novel *TGFBR1* and *TGFBR2* mutations in Marfan syndrome-related disorders. Hum Mutat 27:760–769

Mizuguchi T, Collod-Beroud G, Akiyama T, Abifadel M, Harada N, Morisaki T, Allard D, Varret M, Claustres M, Morisaki H, Ihara M, Kinoshita A, Yoshiura K, Junien C, Kajii T, Jondeau G, Ohta T, Kishino T, Furukawa Y, Nakamura Y, Niikawa N, Boileau C, Matsumoto N (2004) Heterozygous *TGFBR2* mutations in Marfan syndrome. Nat Genet 36:855–860

Neptune ER, Frischmeyer PA, Arking DE, Myers L, Bunton TE, Gayraud B, Ramirez F, Sakai LY, Dietz HC (2003) Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. Nat Genet 33:407–11

Nijbroek G, Sood S, McIntosh I, Francomano CA, Bull E, Pereira L, Ramirez F, Pyeritz RE, Dietz HC (1995) Fifteen novel *FBN1* mutations causing Marfan syndrome detected by heteroduplex analysis of genomic amplicons. Am J Hum Genet 57:8–21

Otto E, Betz R, Rensing C, Schatzle S, Kuntzen T, Vetsi T, Imm A, Hildebrandt F (2000) A deletion distinct from the classical homologous recombination of juvenile nephronophthisis type 1 (NPH1) allows exact molecular definition of deletion breakpoints. Hum Mutat 16:211–223

Pannu H, Fadulu VT, Chang J, Lafont A, Hasham SN, Sparks E, Giampietro PF, Zaleski C, Estrera AL, Safi HJ, Shete S, Willing MC, Raman CS, Milewicz DM (2005) Mutations in transforming growth factor-beta receptor type II cause familial thoracic aortic aneurysms and dissections. Circulation 112:513–520

Pepe G, Giusti B, Evangelisti L, Porciani MC, Brunelli T, Giurlani L, Attanasio M, Fattori R, Bagni C, Comeglio P, Abbate R, Gensini GF (2001) Fibrillin-1 (*FBN1*) gene frameshift mutations in Marfan patients: genotype-phenotype correlation. Clin Genet 59:444–450

Pereira L, Lee SY, Gayraud B, Andrikopoulos K, Shapiro SD, Bunton T, Biery NJ, Dietz HC, Sakai LY, Ramirez F (1999) Pathogenetic sequence for aneurysm revealed in mice underexpressing fibrillin-1. Proc Natl Acad Sci USA 96:3819–3823

Pramparo T, Mattina T, Gimelli S, Liehr T, Zuffardi O (2005) Narrowing the deleted region associated with the 15q21 syndrome. Eur J Med Genet 48:346–352

Qiu P, Soder GJ, Sanfiorenzo VJ, Wang L, Greene JR, Fritz MA, Cai XY (2003) Quantification of single nucleotide polymorphisms by

automated DNA sequencing. Biochem Biophys Res Commun 309:331–338

Rabbee N, Speed TP (2006) A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics 22:7–12

Robinson PN, Arteaga-Solis E, Baldock C, Collod-Beroud G, Booms P, De Paepe A, Dietz HC, Guo G, Handford PA, Judge DP, Kielty CM, Loeys B, Milewicz DM, Ney A, Ramirez F, Reinhardt DP, Tiedemann K, Whiteman P, Godfrey M (2006) The molecular genetics of Marfan syndrome and related disorders. J Med Genet 43:769–787

Rommel K, Karck M, Haverich A, von Kodolitsch Y, Rybczynski M, Muller G, Singh KK, Schmidtke J, Arslan-Kirchner M (2005) Identification of 29 novel and nine recurrent fibrillin-1 (*FBN1*) mutations and genotype-phenotype correlations in 76 patients with Marfan syndrome. Hum Mutat 26:529–539

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30:e57

Schrijver I, Liu W, Odom R, Brenn T, Oefner P, Furthmayr H, Francke U (2002) Premature termination mutations in *FBN1*: Distinct effects on differential allelic expression and on protein and clinical phenotypes. Am J Hum Genet 71:223–237

Shur N, Cowan J, Wheeler PG (2003) Craniosynostosis and congenital heart anomalies associated with a maternal deletion of 15q15-22.1. Am J Med Genet A 120:542–546

Singh KK, Elligsen D, Liersch R, Schubert S, Pabst B, Arslan-Kirchner M, Schmidtke J (2007) Multi-exon out of frame deletion of the *FBN1* gene leading to a severe juvenile onset cardiovascular phenotype in Marfan syndrome. J Mol Cell Cardiol 42:352–356

Contributions of authors to the manuscript "Large Genomic Fibrillin-1 (*FBN1*) Gene Deletions Provide Evidence for True Haploinsufficiency in Marfan Syndrome"

Gábor Mátyás	Conceptual planning, design and supervision of the study, interpretation of data, writing and editing of the manuscript
Sira Alonso	Design of the study, PCR, MLPA, microarray experiments, interpretation of data, writing and editing of the manuscript
Andrea Patrignani	Microarray experiments, editing of the manuscript
Myriam Marti	Microarray experiments, editing of the manuscript
Eliane Arnold	Mutation analyses, editing of the manuscript
István Magyar	RNA analyses, editing of the manuscript
Caroline Henggeler	Mutation analyses, editing of the manuscript
Thierry Carrel	Sending of patients' material and clinical data, editing of the manuscript
Beat Steinmann	Sending of patients' material and clinical data, editing of the manuscript
Wolfgang Berger	Editing of the manuscript

4 General Discussion

4.1 RNA and Disease

Accurate expression and processing of protein-coding and non-coding RNAs is crucial for normal cell functions. RNAs are involved in transcription, RNA processing, protein translation, amino acid transfer, ribosome structure, gene silencing, and imprinting. But they also participate in other processes, where their function has not been clearly defined yet.

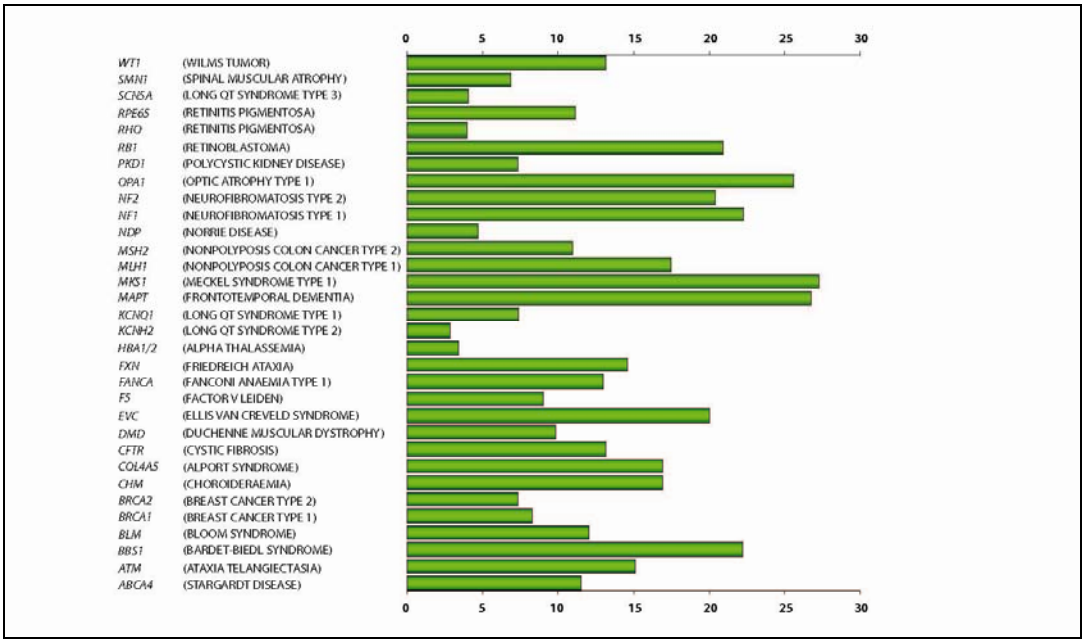

Figure 14. Frequency of splicing defects in common human genetic disorders.
(Data calculated from the Human Gene Mutation Database, HGMD® Professional 2011.2 Release date 24th June 2011.)

Messenger RNAs (mRNAs) are synthesized as precursor RNAs (pre-mRNAs) during the transcription process in the nucleus. The primary transcripts still contain intronic sequences, which are precisely and rapidly removed during the pre-mRNA splicing process. This sophisticated process involves 5 small nuclear ribonucleoproteins (snRNPs) and numerous protein factors, constituting together the engine of splicing, the so called spliceosome. In addition, specific sequence motifs within the pre-mRNA are needed to accurately define exons [Dreyfuss et al., 2002; Glisovic et al., 2008]. Alterations in RNAs (*cis*-acting mutations) and protein factors (*trans*-acting mutations) required for their assembly can be deleterious and disease causing. Indeed, perturbed splicing is linked to homozygous or inactivating mutations in the *SMN1* gene, which cause spinal muscular atrophy (SMA), a

neurodegenerative disease. *SMN1* encodes the SMN protein, which has a role in the maturation of spliceosomal components. Recently it was shown, in SMA mice, that Smn deficiency causes cell type- and snRNP-specific effects. Using exon microarray, the authors have shown that alteration in the stoichiometry of snRNPs resulted in widespread pre-mRNA splicing defects in numerous transcripts of diverse genes [Zhang et al., 2008]. Another interesting example for trans-acting mutations is associated with Prader-Willi syndrome, a congenital disease caused by the loss of paternal gene expression from the maternally imprinted region of chromosome 15q. This region contains 47 almost identical copies of the HBII-52 small nucleolar RNA (snoRNA) gene, which regulates splicing of the serotonin receptor 2C (5-HT_{2C}R). A recent study provided evidence that HBII-52 regulates the processing of 5-HT_{2C}R and the loss of this snoRNA results in an aberrantly spliced mRNA of the serotonin receptor [Kishore et al., 2006]. These few examples indicate how the disruption of the splicing machinery can contribute to disease severity and susceptibility. On the other hand, a substantially large fraction of disease-causing alterations is known to disrupt *cis*-acting elements in the mRNA itself that are important for accurate splicing. The frequency of splicing mutations varies between individual genes. It is likely that up to about 30% of mutations which are linked to human genetic diseases affect pre-mRNA splicing (Figure 14). About half of them disrupt intronic splicing motifs, and the remaining affect exonic sequences, causing missense, nonsense, or even silent mutations, respectively [Nissim-Rafinia and Kerem, 2005; Baralle et al., 2009].

Most of the mutation analyses are performed at the genomic DNA level and the effect of mutations on the encoded mRNA and protein is only predicted based on these sequencing results. However, the effect of the exonic variants can be misclassified, if the analysis is limited to genomic DNA sequence alone. While missense mutations are expected to cause amino acid changes that may be important for structure and function of the protein, silent mutations are not necessarily expected to cause a disease, because the nucleotide change does not lead to an amino acid substitution. Silent mutations are therefore often classified as benign polymorphisms. Frameshift and nonsense mutations are considered to produce transcripts containing premature termination codons, which are degraded by the nonsense-mediated mRNA decay pathway. In fact, however, a significant part of these alterations may have dramatic effects on pre-mRNA processing [Caputi et al., 2002; Cartegni et al., 2002; Baralle and Baralle, 2005]. Moreover, it is also important to identify the consequences of intronic variants located in splicing motifs, if and in which way they cause aberrant splicing,

and to determine the effect of deeper intronic alterations. These sequence variants affect pre-mRNA splicing in different ways: The most common consequence is exon skipping, by which the entire exon will be spliced out. The second frequent effect is cryptic splice site activation, the generation of a novel splice site. As a consequence, exonic or intronic parts of the pre-mRNA are included or excluded from the mature transcript. Rarely, also intron retention, where the entire intron is retained between two exons, and pseudoexon generation, where part of an intron is identified as exon, can be detected. Moreover, any combination of these events can occur.

4.2 *FBN1* Sequence Variants and Aberrant pre-mRNA Splicing

Relatively little is known about *FBN1* pre-mRNA splicing. Only few case reports in the literature are pointing out the importance of *FBN1* transcript analyses (Table 2). A systematic study has not been reported so far. The main issue of this dissertation was an extensive and systematic *in vitro* and *in silico* investigation of exonic and intronic sequence variations in the *FBN1* gene with respect to their effect on splicing (Paper 1). In the first part of this dissertation, I investigated the effect of 125 exonic and intronic mutations at the mRNA level using RNA from different tissues. I compared the results with predicted effects on RNA processing obtained from different *in silico* splice site and/or ESE prediction tools. Overall, ~22% of the sequence variants caused aberrant *FBN1* mRNA processing. Four of them were exonic variants. Of these, three lead to exon skipping and one activates a cryptic splice site (Patient 117, Paper 1). In addition, cryptic splice sites were observed by analyzing 15 intronic mutations (Table 3, Paper 1). Activation of cryptic splice site is the second most frequent splicing defect after exon skipping, as it has been shown in the *OPA1*, *MLH1*, *MSH2*, *BRCA1*, and *BRCA2* genes as well [Bonnet et al., 2008; Schimpf et al., 2008; Arnold et al., 2009]. The question arises: what constitutes a cryptic splice site? Normally, these sites are never selected for splicing even they obviously match to consensus sequences. If those sequences are not found, exon skipping can occur. Alternatively, mRNA folding may play an important role by alteration of the secondary structure of the transcript, allowing de novo consensus sequence and the U1 snRNP to base pairing. It should be mentioned here, that the conservation is just partial and the motifs are degenerate. The human 5` consensus sequence reads MAG|GURAGU (M is A or C, and R is purin) representing the last 3 nucleotides of the exon and the first 6 nucleotides of the intron [Roca et al., 2003]. Accordingly, the 5` splice sites of the human *FBN1* gene follow these specifications (exclusively the donor site of exon 62 because intron 62 is not a GU-AG intron, for more details see the discussion chapter of

Paper 1). It is important to note, that *FBN1* utilizes frequently a U at the flexible position indicated with M (see above and Figure 15, and Paper 1, Supplementary Figure 1).

Table 2. Summary of published *FBN1* mutations affecting pre-mRNA splicing.
(The order of the listings reflect the position within the gene, starting at the 5` end.)

Sequence variant		Affected exon/intron	RNA source	Results of in vitro cDNA analysis	Reference
Nucleotide change	**AA change**				
c.247+1G>A	-	Intron 2	Fibroblasts, Venous blood	Exon 2 skipping	[Dietz et al., 1993a; Guo et al., 2001 Magyar et al., 2009]
c.347+2A>G	-	Intron 3	Fibroblasts, Venous blood	Insertion of 1 bp from intron 3	[Magyar et al., 2009]
c.3208G>C	p.D1070H	Exon 26 (last nucleotide)	Minigene	Insertion of 126 bp from intron 26 Deletion of 47 bp from exon 26	[Chao et al., 2010]
c.4747+5G>T	-	Intron 37	Fibroblasts	Deletion of 48 bp from exon 37	[McGrory and Cole, 1999]
c.5788+1G>A	-	Intron 46	Fibroblasts	Insertion of 33 bp from intron 46	[Hutchinson et al., 2001]
c.6353T>G	p.Y2113X	Exon 51	Fibroblasts	Exon 51 skipping	[Caputi et al., 2002]
c.6354C>T	p.I2118I	Exon 51	Fibroblasts	Exon 51 skipping	[Liu et al., 1997]
c.6379+6T>G	-	Intron 51	Fibroblasts	Insertion of 5 bp from intron 51	[Magyar et al., 2009]
c.6739+1G>C c.6739+3G>T	-	Intron 54	Fibroblasts	Skipping of exon 54	[Godfrey et al., 1993]
c.8051+373G>T	-	Intron 63	Fibroblasts	Pseudoexon, inclusion of 93 bp in intron 63	[Guo et al., 2008]

By comparing 14 cryptic 5` splice site sequences with the normal *FBN1* consensus sequence (Tables 2 and 3, Paper 1), all four nucleotides are tolerated at positions -3 and -2. While the G at position -1 is highly conserved in the normal splice consensus sequence, cryptic splice sites are more tolerant towards T and C (Figure 13). Of the intronic nucleotides, the G at +1 is invariable, regardless whether normal or cryptic sites are considered. The normally invariable T at +2 can occasionally be substituted by a G or A in the cryptic site. The remaining sequences do not display obvious differences between authentic and cryptic 5` splice sites (Figure 15).

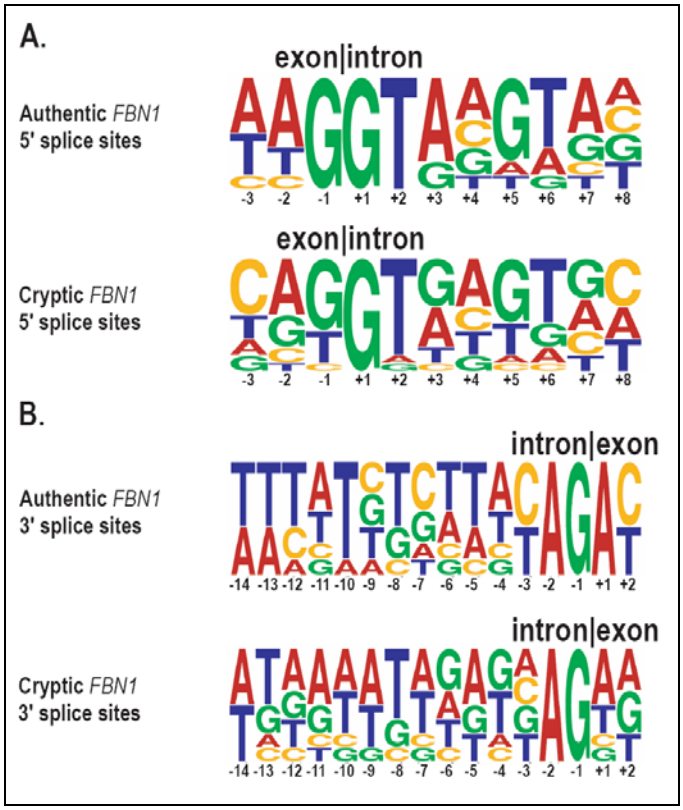

Figure 15. Comparison of authentic and cryptic *FBN1* splice sites.
A. Cryptic 5' splice donor sites generated by different mutations in the *FBN1* gene (Paper 1, Tables 2 and 3).
B. Cryptic 3' splice acceptor sites generated by different mutations in the *FBN1* gene (Paper 1, Table 3).
(The size of a pictogram character is proportional to the frequency of each nucleotide, with the most frequent at the top. Sequence alignments were performed and displayed by using the public online sequence logo generator WebLogo3 http://weblogo.berkeley.edu) [Schneider et al., 1990; Crooks et al., 2004].)

The 3` end of the introns is recognized by three splicing factors, namely the mBBP/SF1 (mammalian branchpoint binding protein/splicing factor 1), the large (65 kDa) and the small (35 kDa) subunits of U2AF (U2 auxiliary factor). mBBP/SF1 binds to the branch point sequence, the U2AF65 interacts with the polypirimidine tract (PPT), while the U2AF35 recognizes the conserved AG dinucleotides at the end of intron. *FBN1* 3` multipart sequences (Figure 15, and Paper 1, Supplementary Figure 1) differ not significantly from consensus sequences derived from other eukaryotic genes. However, one striking difference is a diversion from the highly conserved A at position +1, allowing all four nucleotides at this positions in other eukaryotic genes.

I also compared 8 cryptic acceptor sites (Table 3, Paper 1) with the authentic counterpart based on my *in vitro* results. Interestingly, substantial deviation of the PPT in the cryptic splice sites was observed. Frequency analysis of the 8 cryptic acceptor sites resulted in high purin content within the “polypirimidine” tract. These results suggest a very weak PPT for cryptic splice sites. The AG dinucleotides at the end of introns remained highly conserved,

which may compensate for the weak PPT. Indeed, a previous study provided important insight into the requirement and selection of the PPT. The authors found that a functional PPT does not entirely require uninterrupted uridines, although increased continuous uridine stretches leads to higher efficiency and competitiveness in splicing [Coolidge et al., 1997]. The first two nucleotides of the cryptic *FBN1* exons did not show any conservation. This observation is particularly interesting because under normal conditions aberrant splicing can be expected if the first base of the exon is altered. In the authentic *FBN1* exons, nucleotide A is found with almost invariable frequency, while comparative analysis of different eukaryotic acceptor sites showed that the +1 position can be occupied by any of the four nucleotides (Supplementary Figure1, Paper 1). This observation can explain why in two cases of our patient cohort (Patients 7 and 395, Paper 1, both diverting the A at +1) no aberrant splicing was detected. Similar results have been reported from experiments with minigene constructs, which examined the effect of the *FBN1* c.3209A>G (p.D1070G) mutation on pre-mRNA splicing [Chao et al., 2010]. The authors could show that no aberrantly spliced product was detectable if the first nucleotide of exon 27 was affected.

It appears that in addition to the 5` and 3` splicing signals, also secondary structures within the pre-mRNA play an important role in accurate transcript processing. Indeed, mutations that alter the secondary structure of the RNA can directly inhibit and aid the splicing process. An interesting example was provided by a short deletion in the mouse fibronectin EDA exon (extra domain A exon, one of the alternatively spliced fibronectin exon). This deletion resulted in the shift of a critical ESE from single- into double-stranded structure. This conformational change hindered the SF2/ASF protein to bind and thereby caused the skipping of the affected exon [Buratti and Baralle, 2004]. Further experimental tests by others demonstrated that known splicing motifs preferentially are in a single-stranded context, which provides a stronger effect on splice site selection than double stranded regions. These data indicated that secondary structures contribute to the mRNA splicing process by determining exon recognition [Hiller et al., 2007]. According to this knowledge, I can hypothesize that the here reported cryptic splice sites might be sequestered in structures, and, if the structure is disrupted directly or indirectly by a DNA sequence alteration, the splicing machinery can use these incorrect splice sites.

4.3 Bioinformatical Approaches to Predict the Effect of *FBN1* Sequence Variants on pre-mRNA Splicing

Bioinformatic prediction can be very helpful to predict whether a given sequence variant is a harmless polymorphism or a disease-causing mutation also in the routine genetic diagnostic setting, especially in the absence of RNA samples. These algorithms play an important role in evaluation of the consequence of sequence variants on pre-mRNA splicing. Several computational tools have been developed and published in the last years to predict the effect of exonic substitutions on protein structure and activity (e.g. PolyPhen, SIFT), RNA structure (mfold 3.2), or pre-mRNA splicing by investigating the exonic splicing enhancer (ESE) status (e.g. ESEfinder 3.0, RESCUE-ESE, PESX) [Cartegni et al., 2003; Ng and Henikoff, 2003; Fairbrother et al., 2004; Zhang and Chasin, 2004; Schymkowitz et al., 2005; Zhang et al., 2005].

Table 3. Computational tools for exonic splicing enhancer motif identification used in this study.

Prediction Program	Basic Principle	Input	Output	Website	Reference
ESEfinder 3.0	The prediction of 4 SR protein specific ESE is based on in vitro SELEX approach	Sequences in Fasta format, multiple sequences are allowed	Sequence summary; Tabular output with high scores; Graphical output	http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi	[Cartegni et al., 2003]
RESCUE-ESE	Based on the frequency of hexanucleotide sequences	Maximum 4 kb sequences in Fasta format, multiple Fasta or plain text format are allowed	Yellow hexanucleotide motif(s) above the sequence	http://genes.mit.edu/burgelab/rescue-ese/	[Fairbrother et al., 2004]
PESX	Based on the frequency of octanucleotide sequences, also silencer motifs are predicted	Maximum 10 kb sequences in Fasta or plain text format	Red: silencer motif; Green: enhancer motif; Underlined and bold letter: first bases of enhancer or silencer motifs	http://cubweb.biology.columbia.edu/pesx/	[Zhang and Chasin, 2004; Zhang et al., 2005]

In addition, a myriad of prediction tools are currently available to investigate the effects of intronic variants on pre-mRNA splicing (e.g. NNSplice, NetGene2, Human Splicing Finder, MaxEntScan, ESEfinder 3.0); either as a standalone program or a part of an interface [Brunak et al., 1991; Reese et al., 1997; Cartegni et al., 2003; Yeo and Burge, 2004; Desmet et al., 2009]. In my dissertation, I also evaluated the efficiency of various ESE and splice site

prediction programs for the *FBN1* gene. Details, descriptions, and differences about the applied programs in Paper 1 are summarized here in Tables 4 and 5.

My results show that the use of three commonly available ESE prediction tools (ESEfinder 3.0, RESCUE-ESE, and PESX) was not always reliable for assessing effects of exonic variants. Only ESEfinder 3.0 was able to identify the splicing defect caused by one exonic variant which lies within a putative ESE. The other two programs did not identify true positives but according to specificity and accuracy they appeared more accurate.

These findings are consistent with published data from other comparative studies investigating *MLH1*, *MSH2*, *BRCA1*, and *BRCA2* sequence variants [Arnold et al., 2009; Auclair et al., 2006; Chenevix-Trench et al., 2006]. These data support the notion that these programs should only be used as an indicator. They cannot replace *in vitro* experiments, especially RNA-based investigations. In my opinion, one significant deficiency of these algorithms is the lack of consideration of RNA secondary structure, which may be influenced by mutations. This point is particularly important because changes in the nucleotide sequence could have considerable consequences on RNA structure, as I discussed above. The present state of ESE predicton tools is such that their usefulness applies primarily to the interpretation of experimental data rather than simple predictions (personal communication with A. Krainer and L. Cartegni, developers of the ESEfinder prediction program).

For *in silico* prediction of intronic variants five bioinformatic tools were chosen, the Human Splicing Finder version 2.4 (HSF), the MaxEntScan, the NNSPLICE version 0.9, the NetGene2 splice site prediction tools, and the splice site analysis function of the previously introduced ESEfinder 3.0. All five programs provided good predictions with HSF, MES, and NetGene2 scoring best. I have to emphasize that it would be desirable to have prediction programs that also could indicate the potential consequence of an intronic mutation (e.g. cryptic splice sites, or complex splicing events). HSF has this ability but predictions could rarely be confirmed with *in vitro* experiments.

It is important to mention that these programs are also not validated for use in diagnostics. In the present state they cannot replace *in vitro* studies but may help to prioritize samples for *in vitro* analysis [Arnold et al., 2009].

Table 4. Computational tools for splice site assessment and identification used in this study.

Prediction Program	Basic Principle	Input	Output	Website	Reference
Human Splicing Finder 2.4	The donor and acceptor site predictions based on position weight matrices	Maximum 2.5 kb sequences in plain text format	14 nt motifs for acceptor site and 9 nt motifs for donor site with numerical score (0-100)	http://www.umd.be/SSF/	[Desmet et al., 2009]
MaxEntScan	The program based on the use of different scoring models	9-mer sequences for 5' ss and 23-mer for 3' ss in Fasta format; Maximum 2.5 kb sequences in plain text format in the HSF interface	23 nt motifs for acceptor site and 9 nt motifs for donor site with numerical score (maximal scores: 11.81 for 5` ss and 13.59 for 3` ss)	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html http://www.umd.be/SSF/	[Yeo and Burge, 2004; Desmet et al., 2009]
NNSplice	Based on backpropagation feedforward neural network with one layer of hidden units to recognize donor and acceptor sites	Sequences or multiple sequences in Fasta format	41 nt motifs for acceptor site and 15 nt motifs for donor site with numerical score above the threshold (0.4-1)	http://www.fruitfly.org/seq_tools/splice.html	[Reese et al., 1997]
NetGene2	Uses a neural network combined with rule-based system to predict splice sites	Maximum 80 kb sequences in Fasta format	20 nt motifs for acceptor site and 20 nt motifs for donor site with numerical score (0-1)	http://www.cbs.dtu.dk/services/NetGene2/	[Brunak et al., 1991]
ESEfinder 3.0	The program uses different thresholds for mouse and human splice sites	Maximum 5 kb sequences in Fasta or multiple Fasta format are allowed	30 nt motifs for acceptor site and 30 nt motifs for donor site with high scores above the threshold	http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi	[Cartegni et al., 2003]

4.4 *FBN1* and Nonsense-mediated mRNA Decay

In mammalian cells, NMD is responsible for the degradation of premature termination codon (PTC) harboring transcripts. PTC-bearing transcripts cannot only arise from frameshift and nonsense mutations but can also be a result of splicing or transcription errors. The

elimination of such aberrant products is particularly important, since they can translate into proteins with dominant-negative or deleterious gain-of-function effects [Gudikote et al., 2005]. The degradation process depends on both translation and pre-mRNA splicing [Chan et al., 2007].

In Paper 2, we described our recent observation that NMD is incomplete in leukocytes, while in fibroblast cell lines PTC-containing *FBN1* transcripts undergo an efficient degradation process. In developing lymphocytes, the process of degradation of NMD-sensitive transcripts is essential. During this developmental process, programmed DNA rearrangements of gene sequences encoding T-cell receptors and immunoglobulins is necessary to generate a diverse set of receptors which recognize different antigens. Approximately two third of these rearrangement result in transcripts containing PTCs. Interestingly, transcripts generated from re-arranged genes undergo more efficient downregulation (~1-5% of normal levels) compared to those from non-rearranged genes (~10-30% of normal levels). Specifically, a necessary and sufficient regulatory sequence for NMD has been localized to a region composed of the VDJ exon and flanking introns of the T-cell receptor-β (TCRβ) genes [Gudikote and Wilkinson, 2002]. We have not investigated NMD directly in fibroblast cell lines and blood but different regulatory sequences in fibroblasts and leukocytes could explain our results. The role of the 5` upstream region of the *FBN1* gene is relatively unknown. It contains 3 alternative 5` exons (exons B, A, and C) which can be spliced individually to the first constitutive coding exon. One can speculate whether such regulatory sequences, which were found in the TCRβ genes, may also occur in the *FBN1* 5` region. Alternatively, tissue-specific NMD pathways can exist in fibroblasts and leukocytes. The first hint about tissue-specific NMD was discussed in the study published by Bateman and colleagues on mutations leading to PTC in the collagen X gene (*COL10A1*) [Bateman et al., 2003]. They could show that NMD completely triggered the degradation of PTC-containing *COL10A1* mutant transcripts in cartilage tissue but not in non-cartilage cells. Further evidence was delivered by murine studies where in 13 different tissues the relative amount of PTC-containing *Men1* transcripts were measured. The authors found up to two-fold difference in the efficiency of NMD between the examined tissues [Zetoune et al., 2008]. Moreover, in HeLa and MCF7 cell lines, Linde and her colleagues could demonstrate the existence of cells in which NMD of all transcripts is efficient, whereas in other cells NMD was less efficient [Linde et al., 2007]. These data suggest that the efficiency of the RNA surveillance is cell-specific. The results of these studies suggest the existence of distinct

branches of the NMD pathway. Indeed, to date three branches of the NMD pathway have been described in mammals [Chang et al., 2007]. These alternative branches can be discriminated based on NMD factors, the exon-junction complex, and cell-type specific co-factors. It remains for future studies to investigate whether different pathways and factors operate for PTC-containing *FBN1* transcripts in blood and fibroblasts.

Incomplete NMD in leukocytes could have important implication in *FBN1* molecular genetic testing, allowing RNA-based mutation detection in blood and offering efficient screening of the exons of other large genes. Moreover, it can be used to identify aberrantly spliced transcripts as well as transcripts with frameshift mutations, which would remain undetectable by standard genomic DNA or complementary DNA screening if RNA was harvested from cells grown without protein synthesis inhibitor (e.g. cycloheximide, anisomycin, emetine, puromycin). However, we have to be careful in interpretation of data in diagnostically investigated materials. For correct conclusion regarding the expression of the NMD-sensitive transcript, it is important to confirm the level of NMD in the affected tissue.

4.5 Quantitative Sequencing for Transcript Analyses

For correct interpretation of genotype-phenotype correlations, the determination of allelic expression and consequences of a given nucleotide substitution at the level of RNA is essential. To quantify the mutant versus wild-type transcript level we improved, evaluated, and applied a quantitative Sanger sequencing method. Using this technology we were able to measure the relative level of mutant versus wild-type transcripts for 70 *FBN1* exonic variants (Paper 1) and 26 PTC mutations (Paper 2).

Investigation of allele-specific expression is particularly important. All heterozygous sequence alterations in the coding region can be used as markers to determine the relative amount of allele-specific expression [Hartmann et al., 2008]. If the mutation does not interfere with allelic expression, we would expect to detect a 1:1 ratio from each allele. This was observed for example when we investigated potential alterations for an *SLC16A12* mutation (Appendix 2, Paper 5). In contrast, non-equal allele representation in the mRNA transcripts (divergence of a 1:1 allelic expression ratio), which can lead to a complete loss of transcripts from one allele, could indicate aberrant splicing, NMD-based degradation, or different expression activity of the two alleles. However, we occasionally observed different amounts of *FBN1* transcripts from the two alleles, even if no aberrant splicing was detectable. Similarly, the investigation of allele-specific expression of five other genes carrying

heterozygous mutations were reported in the literature [Yan et al., 2002]. In addition, intrafamilial allelic variation was found for *Calpain-10* and *PKD2* expression. Such variation was also observed by investigating the allele-specific expression of the *RPGR* gene in two unrelated families (Appendix 1, Paper 4).

Clinical variability of Marfan syndrome in unrelated individuals or family members carrying the same *FBN1* mutation could be explained by this phenomenon [Hutchinson et al., 2003].

4.6 Towards Understanding of Marfan Syndrome

So far, there is no cure for Marfan syndrome but medical and surgical treatment and management can improve and extend the life expectancy of Marfan patients [Dean, 2007]. Recently, clinical trials have been started in order to treat patients with angiotensin II type 1 receptor blocker (losartan), which is frequently used to treat hypertension. In a preliminary study, it was shown that losartan not only lowers blood pressure but can also decrease the rate of progressive dilatation of the aorta. Interestingly, variability in the response to the therapy has previously been observed. This could be caused by a multitude of factors such as genotype, age, body mass, and/or gender [Brooke et al., 2008]. Several examples are known that demonstrate the influence of genetic factors on the outcome of drug treatment. For example, Kloeckener-Gruissem and her colleagues reported that an individual`s response to intravitreal ranibizumab application for treatment of neovascular age-related macular degeneration can be influenced by genetic factors which contribute to a positive or negative outcome of treatment [Kloeckener-Gruissem et al., 2011]. Furthermore, response to interferon-alpha and ribavirin, a treatment of chronic hepatitis C, appears to be dependent on the patient`s genotype [Hwang et al., 2006]. Thanks to novel technological innovations, such as next-generation sequencing, the determination of the individual genetic profile could be routinely available in the near future. This technology dramatically increases the speed and reduces the costs of genetic analyses. Hopefully, detailed genome-wide analyses of Marfan patients will become more readily available so that the genetic information will be no more a limiting factor in application of genetics to clinical medicine. The integration of data from these two fields will assist in the clinical decision making process and contribute significantly towards personalized medicine [Ashley et al., 2010]. In this data accumulation, one of the major challenges is to understand the role and pathogenic mechanism of particular sequence variants and to discriminate benign polymorphisms and pathophysiologically relevant

sequence variations. Hopefully, the data and results of the present dissertation will contribute and make this understanding easier.

5 References

Achard EC. 1902. Arachnodactylie. Bull Mem Soc Med Hop Paris 19:834-843.

Ahram D, Sato TS, Kohilan A, Tayeh M, Chen S, Leal S, Al-Salem M, El-Shanti H. 2009. A homozygous mutation in ADAMTSL4 causes autosomal-recessive isolated ectopia lentis. Am J Hum Genet 84(2):274-278.

Altman A, Uliel L, Caspi L. 2008. Dural ectasia as presenting symptom of Marfan syndrome. Isr Med Assoc J 10(3):194-195.

Aoyama T, Tynan K, Dietz HC, Francke U, Furthmayr H. 1993. Missense mutations impair intracellular processing of fibrillin and microfibril assembly in Marfan syndrome. Hum Mol Genet 2(12):2135-2140.

Aoyama T, Francke U, Dietz HC, Furthmayr H. 1994. Quantitative differences in biosynthesis and extracellular deposition of fibrillin in cultured fibroblasts distinguish five groups of Marfan syndrome patients and suggest distinct pathogenetic mechanisms. J Clin Invest 94(1):130-137.

Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, Woods MO, Hopper JL, Jenkins MA, Brown MA, Tavtigian SV, Goldgar DE, Young JP, Spurdle AB. 2009. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. Hum Mutat 30(5):757-770.

Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, Pushkarev D, Neff NF, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Knowles JW, Chou MF, Thakuria JV, Rosenbaum AM, Zaranek AW, Church GM, Greely HT, Quake SR, Altman RB. 2010. Clinical assessment incorporating a personal genome. Lancet 375(9725):1525-1535.

Auclair J, Busine MP, Navarro C, Ruano E, Montmain G, Desseigne F, Saurin JC, Lasset C, Bonadona V, Giraud S, Puisieux A, Wang, Q. 2006. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. Hum Mutat 27(2):145-154.

Baralle D, Baralle M. 2005. Splicing in action: assessing disease causing sequence changes. J Med Genet 42(10):737-748.

Baralle D, Lucassen A, Buratti E. 2009. Missed threads. The impact of pre-mRNA splicing defects on clinical practice. EMBO Rep 10(8):810-816.

Bateman JF, Freddi S, Nattrass G, Savarirayan R. 2003. Tissue-specific RNA surveillance? Nonsense-mediated mRNA decay causes collagen X haploinsufficiency in Schmid metaphyseal chondrodysplasia cartilage. Hum Mol Genet 12(3):217-225.

Beighton P, de Paepe A, Danks D, Finidori G, Gedde-Dahl T, Goodman R, Hall JG, Hollister DW, Horton W, McKusick VA, Opitz JM, Pope FM, Pyeritz RE, Rimoin DL, Sillence D, Spranger JW, Thompson E, Tsipouras P, Viljoen D, Winship I, Young I, Reynolds JF. 1988. International Nosology of Heritable Disorders of Connective Tissue, Berlin, 1986. Am J Med Genet 29(3):581-594.

Biery NJ, Eldadah ZA, Moore CS, Stetten G, Spencer F, Dietz HC. 1999. Revised genomic organization of FBN1 and significance for regulated gene expression. Genomics 56(1):70-77.

Blencowe BJ. 2000. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci 25(3):106-110.

Bonnet C, Krieger S, Vezain M, Rousselin A, Tournier I, Martins A, Berthet P, Chevrier A, Dugast C, Layet V, Rossi A, Lidereau R, Frébourg T, Hardouin A, Tosi M. 2008. Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription-PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. J Med Genet 45(7):438-446.

Brooke BS, Habashi JP, Judge DP, Patel N, Loeys B, Dietz HC, 3rd. 2008. Angiotensin II blockade and aortic-root dilation in Marfan's syndrome. N Engl J Med 358(26):2787-2795.

Brunak S, Engelbrecht J, Knudsen S. 1991. Prediction of human mRNA donor and acceptor sites from the DNA sequence. J Mol Biol 220(1):49-65.

Buratti E, Baralle FE. 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. Mol Cell Biol 24(24):10505-10514.

Byers PH. 2004. Determination of the molecular basis of Marfan syndrome: a growth industry. J Clin Invest 114(2):161-163.

Caputi M, Kendzior RJ, Jr., Beemon KL. 2002. A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer. Genes Dev 16(14):1754-1759.

Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3(4):285-298.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 31(13):3568-3571.

Chan WK, Huang L, Gudikote JP, Chang YF, Imam JS, MacLean JA, 2nd, Wilkinson MF. 2007. An alternative branch of the nonsense-mediated decay pathway. Embo J 26(7):1820-1830.

Chang YF, Imam JS, Wilkinson MF. 2007. The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76:51-74.

Chaudhry SS, Cain SA, Morgan A, Dallas SL, Shuttleworth CA, Kielty CM. 2007. Fibrillin-1 regulates the bioavailability of TGFbeta1. J Cell Biol 176(3):355-367.

Chao SC, Chen JS, Tsai CH, Lin JM, Lin YJ, Sun HS. 2010. Novel exon nucleotide substitution at the splice junction causes a neonatal Marfan syndrome. Clin Genet 77(5):453-463.

Chenevix-Trench G, Healey S, Lakhani S, Waring P, Cummings M, Brinkworth R, Deffenbaugh AM, Burbidge LA, Pruss D, Judkins T, Scholl T, Bekessy A, Marsh A, Lovelock P, Wong M, Tesoriero A, Renard H, Southey M, Hopper JL, Yannoukakos K, Brown M, Easton D, Tavtigian SV, Goldgar D, Spurdle AB; kConFab Investigators. 2006. Genetic and histopathologic evaluation of BRCA1 and BRCA2 DNA sequence variants of unknown clinical significance. Cancer Res 66(4):2019-2027.

Collod G, Babron MC, Jondeau G, Coulon M, Weissenbach J, Dubourg O, Bourdarias JP, Bonaiti-Pellie C, Junien C, Boileau C. 1994. A second locus for Marfan syndrome maps to chromosome 3p24.2-p25. Nat Genet 8(3):264-268.

Collod-Beroud G, Boileau C. 2002. Marfan syndrome in the third Millennium. Eur J Hum Genet 10(11):673-681.

Coolidge CJ, Seely RJ, Patton JG. 1997. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. Nucleic Acids Res 25(4):888-896.

Corson GM, Chalberg SC, Dietz HC, Charbonneau NL, Sakai LY. 1993. Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure and alternatively spliced exons at the 5' end. Genomics 17(2):476-484.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14(6):1188-1190.

De Paepe A, Devereux RB, Dietz HC, Hennekam RC, Pyeritz RE. 1996. Revised diagnostic criteria for the Marfan syndrome. Am J Med Genet 62(4):417-426.

Dean JC. 2007. Marfan syndrome: clinical diagnosis and management. Eur J Hum Genet 15(7):724-733.

Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 37(9):e67.

Dietz HC, Cutting GR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, Puffenberger EG, Hamosh A, Nanthakumar EJ, Curristin SM, Stetten G, Meyers DA, Francomano CA. 1991a. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. Nature 352(6333):337-339.

Dietz HC, Pyeritz RE, Hall BD, Cadle RG, Hamosh A, Schwartz J, Meyers DA, Francomano CA. 1991b. The Marfan syndrome locus: confirmation of assignment to chromosome 15 and identification of tightly linked markers at 15q15-q21.3. Genomics 9(2):355-361.

Dietz HC, McIntosh I, Sakai LY, Corson GM, Chalberg SC, Pyeritz RE, Francomano CA. 1993. Four novel FBN1 mutations: significance for mutant transcript level and EGF-like domain calcium binding in the pathogenesis of Marfan syndrome. Genomics 17(2):468-475.

Dreyfuss G, Kim VN, Kataoka N. 2002. Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol 3(3):195-205.

Eldadah ZA, Brenn T, Furthmayr H, Dietz HC. 1995. Expression of a mutant human fibrillin allele upon a normal human or murine genetic background recapitulates a Marfan cellular phenotype. J Clin Invest 95(2):874-880.

Fairbrother WG, Yeo GW, Yeh R, Goldstein P, Mawson M, Sharp PA, Burge CB. 2004. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. Nucleic Acids Res 32(Web Server issue):W187-190.

Faivre L, Dollfus H, Lyonnet S, Alembik Y, Megarbane A, Samples J, Gorlin RJ, Alswaid A, Feingold J, Le Merrer M, Munnich A, Cormier-Daire V. 2003a. Clinical homogeneity and genetic heterogeneity in Weill-Marchesani syndrome. Am J Med Genet A 123A(2):204-207.

Faivre L, Gorlin RJ, Wirtz MK, Godfrey M, Dagoneau N, Samples JR, Le Merrer M, Collod-Beroud G, Boileau C, Munnich A, Cormier-Daire V. 2003b. In frame fibrillin-1 gene deletion in autosomal dominant Weill-Marchesani syndrome. J Med Genet 40(1):34-36.

Faivre L, Collod-Beroud G, Loeys BL, Child A, Binquet C, Gautier E, Callewaert B, Arbustini E, Mayer K, Arslan-Kirchner M, Kiotsekoglou A, Comeglio P, Marziliano N, Dietz HC, Halliday D, Beroud C, Bonithon-Kopp C, Claustres M, Muti C, Plauchu

H, Robinson PN, Adès LC, Biggin A, Benetts B, Brett M, Holman KJ, De Backer J, Coucke P, Francke U, De Paepe A, Jondeau G, Boileau C. 2007. Effect of mutation type and location on clinical outcome in 1,013 probands with Marfan syndrome or related phenotypes and FBN1 mutations: an international study. Am J Hum Genet 81(3):454-466.

Faivre L, Khau Van Kien P, Callier P, Ruiz-Pallares N, Baudoin C, Plancke A, Wolf JE, Thauvin-Robinet C, Durand E, Minot D, Dulieu V, Metaizeau JD, Leheup B, Coron F, Bidot S, Huet F, Jondeau G, Boileau C, Claustres M, Mugneret F. 2010. De novo 15q21.1q21.2 deletion identified through FBN1 MLPA and refined by 244K array-CGH in a female teenager with incomplete Marfan syndrome. Eur J Med Genet 53(4):208-212.

Faivre L, Collod-Beroud G, Adès L, Arbustini E, Child A, Callewaert B, Loeys B, Binquet C, Gautier E, Mayer K, Arslan-Kirchner M, Grasso M, Beroud C, Hamroun D, Bonithon-Kopp C, Plauchu H, Robinson P, De Backer J, Coucke P, Francke U, Bouchot O, Wolf J, Stheneur C, Hanna N, Detaint D, De Paepe A, Boileau C, Jondeau G. 2011. The new Ghent criteria for Marfan syndrome: what do they change? Clin Genet [Epub ahead of print].

Fu XD. 1995. The superfamily of arginine/serine-rich splicing factors. RNA 1(7):663-80.

Giltay R, Timpl R, Kostka G. 1999. Sequence, recombinant expression and tissue localization of two novel extracellular matrix proteins, fibulin-3 and fibulin-4. Matrix Biol 18(5):469-480.

Glisovic T, Bachorik JL, Yong J, Dreyfuss G. 2008. RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett 582(14):1977-1986.

Godfrey M, Vandemark N, Wang M, Velinov M, Wargowski D, Tsipouras P, Han J, Becker J, Robertson W, Droste S, Rao VH. 1993. Prenatal diagnosis and a donor splice site mutation in fibrillin in a family with Marfan syndrome. Am J Hum Genet 53(2):472-480.

Gudikote JP, Wilkinson MF. 2002. T-cell receptor sequences that elicit strong down-regulation of premature termination codon-bearing transcripts. EMBO J 21(1-2):125-134.

Gudikote JP, Imam JS, Garcia RF, Wilkinson MF. 2005. RNA splicing promotes translation and RNA surveillance. Nat Struct Mol Biol 12(9):801-809.

Guo D, Hasham S, Kuang SQ, Vaughan CJ, Boerwinkle E, Chen H, Abuelo D, Dietz HC, Basson CT, Shete SS, Milewicz DM. 2001. Familial thoracic aortic aneurysms and dissections: genetic heterogeneity with a major locus mapping to 5q13-14. Circulation 103(20):2461-2468.

Guo DC, Gupta P, Tran-Fadulu V, Guidry TV, Leduc MS, Schaefer FV, Milewicz DM. 2008. An FBN1 pseudoexon mutation in a patient with Marfan syndrome: confirmation of cryptic mutations leading to disease. J Hum Genet 53(11-12):1007-1011.

Guo DC, Papke CL, Tran-Fadulu V, Regalado ES, Avidan N, Johnson RJ, Kim DH, Pannu H, Willing MC, Sparks E, Pyeritz RE, Singh MN, Dalman RL, Grotta JC, Marian AJ, Boerwinkle EA, Frazier LQ, LeMaire SA, Coselli JS, Estrera AL, Safi HJ, Veeraraghavan S, Muzny DM, Wheeler DA, Willerson JT, Yu RK, Shete SS, Scherer SE, Raman CS, Buja LM, Milewicz DM. 2009. Mutations in smooth muscle alpha-actin (ACTA2) cause coronary artery disease, stroke, and Moyamoya disease, along with thoracic aortic disease. Am J Hum Genet 84(5):617-627.

Guo G, Bauer S, Hecht J, Schulz MH, Busche A, Robinson PN. 2008. A short ultraconserved sequence drives transcription from an alternate FBN1 promoter. Int J Biochem Cell Biol 40(4):638-650.

Halliday D, Hutchinson S, Kettle S, Firth H, Wordsworth P, Handford PA. 1999. Molecular analysis of eight mutations in FBN1. Hum Genet 105(6):587-597.

Hartmann L, Theiss S, Niederacher D, Schaal H. 2008.Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? Front Biosci 13:3252-3272.

Hasham SN, Willing MC, Guo DC, Muilenburg A, He R, Tran VT, Scherer SE, Shete SS, Milewicz DM. 2003. Mapping a locus for familial thoracic aortic aneurysms and dissections (TAAD2) to 3p24-25. Circulation 107(25):3184-3190.

Hecht F, Beals RK. 1972. "New" syndrome of congenital contractural arachnodactyly originally described by Marfan in 1896. Pediatrics 49(4):574-579.

Hilhorst-Hofstee Y, Hamel BC, Verheij JB, Rijlaarsdam ME, Mancini GM, Cobben JM, Giroth C, Ruivenkamp CA, Hansson KB, Timmermans J, Moll HA, Breuning MH, Pals G. 2011. The clinical spectrum of complete FBN1 allele deletions. Eur J Hum Genet 19(3):247-252.

Hiller M, Zhang Z, Backofen R, Stamm S. 2007. Pre-mRNA secondary structures influence exon recognition. PLoS Genet 3(11):e204.

Holbrook JA, Neu-Yilik G, Hentze MW, Kulozik AE. 2004. Nonsense-mediated decay approaches the clinic. Nat Genet 36(8):801-808.

Hutchinson S, Wordsworth BP, Handford PA. 2001. Marfan syndrome caused by a mutation in FBN1 that gives rise to cryptic splicing and a 33 nucleotide insertion in the coding sequence. Hum Genet 109(4):416-420.

Hutchinson S, Furger A, Halliday D, Judge DP, Jefferson A, Dietz HC, Firth H, Handford PA. 2003. Allelic variation in normal human FBN1 expression in a family with Marfan syndrome: a potential modifier of phenotype? Hum Mol Genet 12(18):2269-2276.

Hwang Y, Chen EY, Gu ZJ, Chuang WL, Yu ML, Lai MY, Chao YC, Lee CM, Wang JH, Dai CY, Shian-Jy Bey M, Liao YT, Chen PJ, Chen DS. 2006. Genetic predisposition of responsiveness to therapy for chronic hepatitis C. Pharmacogenomics 7(5):697-709.

Judge DP, Biery NJ, Keene DR, Geubtner J, Myers L, Huso DL, Sakai LY, Dietz HC. 2004. Evidence for a critical contribution of haploinsufficiency in the complex pathogenesis of Marfan syndrome. J Clin Invest 114(2):172-181.

Judge DP, Dietz HC. 2005. Marfan's syndrome. Lancet 366(9501):1965-1976.

Kaartinen V, Warburton D. 2003. Fibrillin controls TGF-beta activation. Nat Genet 33(3):331-332.

Kainulainen K, Pulkkinen L, Savolainen A, Kaitila I, Peltonen L. 1990. Location on chromosome 15 of the gene defect causing Marfan syndrome. N Engl J Med 323(14):935-939.

Kainulainen K, Sakai LY, Child A, Pope FM, Puhakka L, Ryhanen L, Palotie A, Kaitila I, Peltonen L. 1992. Two mutations in Marfan syndrome resulting in truncated fibrillin polypeptides. Proc Natl Acad Sci USA 89(13):5917-5921.

Kishore S, Stamm S, 2006. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. Science 311(5758):230-232.

Kloeckener-Gruissem B, Barthelmes D, Labs S, Schindler C, Kurz-Levin M, Michels S, Fleischhauer J, Berger W, Sutter F, Menghini M. 2011. Genetic association with response to intravitreal ranibizumab (Lucentis(R)) in neovascular AMD patients. Invest Ophthalmol Vis Sci 52(7):4694-4702.

Kosaki K, Takahashi D, Udaka T, Kosaki R, Matsumoto M, Ibe S, Isobe T, Tanaka Y, Takahashi T. 2006. Molecular pathology of Shprintzen-Goldberg syndrome. Am J Med Genet A 140(1):104-8; author reply 109-110.

Krawczak M, Reiss J, Cooper DN. 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 90(1-2):41-54.

Levine RA, Slaugenhaupt SA. 2007. Molecular genetics of mitral valve prolapse. Curr Opin Cardiol 22(3):171-175.

Linde L, Boelz S, Neu-Yilik G, Kulozik AE, Kerem B. 2007. The efficiency of nonsense-mediated mRNA decay is an inherent character and varies among different cells. Eur J Hum Genet 15(11):1156-1162.

Liu W, Qian C, Francke U. 1997. Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. Nat Genet 16(4):328-329.

Liu W, Schrijver I, Brenn T, Furthmayr H, Francke U. 2001. Multi-exon deletions of the FBN1 gene in Marfan syndrome. BMC Med Genet 2(1):11.

Loeys B, De Backer J, Van Acker P, Wettinck K, Pals G, Nuytinck L, Coucke P, De Paepe A. 2004. Comprehensive molecular screening of the FBN1 gene favors locus homogeneity of classical Marfan syndrome. Hum Mutat 24(2):140-146.

Loeys BL, Chen J, Neptune ER, Judge DP, Podowski M, Holm T, Meyers J, Leitch CC, Katsanis N, Sharifi N, Xu FL, Myers LA, Spevak PJ, Cameron DE, De Backer J, Hellemans J, Chen Y, Davis EC, Webb CL, Kress W, Coucke P, Rifkin DB, De Paepe AM, Dietz HC. 2005. A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. Nat Genet 37(3):275-281.

Loeys BL, Schwarze U, Holm T, Callewaert BL, Thomas GH, Pannu H, De Backer JF, Oswald GL, Symoens S, Manouvrier S, Roberts AE, Faravelli F, Greco MA, Pyeritz RE, Milewicz DM, Coucke PJ, Cameron DE, Braverman AC, Byers PH, De Paepe AM, Dietz HC. 2006. Aneurysm syndromes caused by mutations in the TGF-beta receptor. N Engl J Med 355(8):788-798.

Loeys BL, Dietz HC, Braverman AC, Callewaert BL, De Backer J, Devereux RB, Hilhorst-Hofstee Y, Jondeau G, Faivre L, Milewicz DM, Pyeritz RE, Sponseller PD, Wordsworth P, De Paepe AM. 2010. The revised Ghent nosology for the Marfan syndrome. J Med Genet 47(7):476-485.

Lönnqvist L, Reinhardt D, Sakai L, Peltonen L. 1998. Evidence for furin-type activity-mediated C-terminal processing of profibrillin-1 and interference in the processing by certain mutations. Hum Mol Genet 7(13):2039-2044.

Magyar I, Colman D, Arnold E, Baumgartner D, Bottani A, Fokstuen S, Addor MC, Berger W, Carrel T, Steinmann B, Mátyás G. 2009. Quantitative sequence analysis of FBN1 premature termination codons provides evidence for incomplete NMD in leukocytes. Hum Mutat 30(9):1355-1364.

Marfan, A. 1896. Un cas de deformation congenitale des quatre membres, plus prononcee aux extremites, caracterisee par l´allongement des os avec un certain degre d´amincissement. Bull Mem Soc Med Hop Paris 13:220–226.

Massague J, Gomis RR. 2006. The logic of TGFbeta signaling. FEBS Lett 580(12):2811-2820.

Matlin AJ, Clark F, Smith CW. 2005. Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol 6(5):386-398.

Mátyás G, Arnold E, Carrel T, Baumgartner D, Boileau C, Berger W, Steinmann B. 2006. Identification and in silico analyses of novel TGFBR1 and TGFBR2 mutations in Marfan syndrome-related disorders. Hum Mutat 27(8):760-769.

Mátyás G, Alonso S, Patrignani A, Marti M, Arnold E, Magyar I, Henggeler C, Carrel T, Steinmann B, Berger W. 2007. Large genomic fibrillin-1 (FBN1) gene deletions provide evidence for true haploinsufficiency in Marfan syndrome. Hum Genet 122(1):23-32.

McGrory J, Cole WG. 1999. Alternative splicing of exon 37 of FBN1 deletes part of an 'eight-cysteine' domain resulting in the Marfan syndrome. Clin Genet 55(2):118-121.

Mecham RP, Heuser J. 1990. Three-dimensional organization of extracellular matrix in elastic cartilage as viewed by quick freeze, deep etch electron microscopy. Connect Tissue Res 24(2):83-93.

Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. 2004. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet 36(10):1073-1078.

Milewicz DM, Pyeritz RE, Crawford ES, Byers PH. 1992. Marfan syndrome: defective synthesis, secretion, and extracellular matrix formation of fibrillin by cultured dermal fibroblasts. J Clin Invest 89(1):79-86.

Milewicz DM, Dietz HC, Miller DC. 2005. Treatment of aortic disease in patients with Marfan syndrome. Circulation 111(11):e150-157.

Mizuguchi T, Collod-Beroud G, Akiyama T, Abifadel M, Harada N, Morisaki T, Allard D, Varret M, Claustres M, Morisaki H, Ihara M, Kinoshita A, Yoshiura K, Junien C, Kajii T, Jondeau G, Ohta T, Kishino T, Furukawa Y, Nakamura Y, Niikawa N, Boileau C, Matsumoto N. 2004. Heterozygous TGFBR2 mutations in Marfan syndrome. Nat Genet 36(8):855-860.

Moat SJ, Bao L, Fowler B, Bonham JR, Walter JH, Kraus JP. 2004. The molecular basis of cystathionine beta-synthase (CBS) deficiency in UK and US patients with homocystinuria. Hum Mutat 23(2):206.

Morisaki H, Akutsu K, Ogino H, Kondo N, Yamanaka I, Tsutsumi Y, Yoshimuta T, Okajima T, Matsuda H, Minatoya K, Sasaki H, Tanaka H, Ishibashi-Ueda H, Morisaki T. 2009. Mutation of ACTA2 gene as an important cause of familial and nonfamilial nonsyndromatic thoracic aortic aneurysm and/or dissection (TAAD). Hum Mutat 30(10):1406-1411.

Nahum Y, Spierer A. 2008. Ocular features of Marfan syndrome: diagnosis and management. Isr Med Assoc J 10(3):179-181.

Neptune ER, Frischmeyer PA, Arking DE, Myers L, Bunton TE, Gayraud B, Ramirez F, Sakai LY, Dietz HC. 2003. Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. Nat Genet 33(3):407-411.

Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31(13):3812-3814.

Nishimura A, Sakai H, Ikegawa S, Kitoh H, Haga N, Ishikiriyama S, Nagai T, Takada F, Ohata T, Tanaka F, Kamasaki H, Saitsu H, Mizuguchi T, Matsumoto N. 2007. FBN2,

FBN1, TGFBR1, and TGFBR2 analyses in congenital contractural arachnodactyly. Am J Med Genet A 143(7):694-698.

Nissim-Rafinia M, Kerem B. 2005. The splicing machinery is a genetic modifier of disease severity. Trends Genet 21(9):480-483.

Pannu H, Fadulu VT, Chang J, Lafont A, Hasham SN, Sparks E, Giampietro PF, Zaleski C, Estrera AL, Safi HJ, Shete S, Willing MC, Raman CS, Milewicz DM. 2005. Mutations in transforming growth factor-beta receptor type II cause familial thoracic aortic aneurysms and dissections. Circulation 112(4):513-520.

Pereira L, Andrikopoulos K, Tian J, Lee SY, Keene DR, Ono R, Reinhardt DP, Sakai LY, Biery NJ, Bunton T, Dietz HC, Ramirez F. 1997. Targetting of the gene encoding fibrillin-1 recapitulates the vascular aspect of Marfan syndrome. Nat Genet 17(2):218-222.

Pereira L, D'Alessio M, Ramirez F, Lynch JR, Sykes B, Pangilinan T, Bonadio J. 1993. Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan syndrome [published erratum appears in Hum Mol Genet 1993 Oct;2(10):1762]. Hum Mol Genet 2(7):961-968.

Putnam EA, Zhang H, Ramirez F, Milewicz DM. 1995. Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder, congenital contractural arachnodactyly. Nat Genet 11(4):456-458.

Raghunath M, Kielty CM, Kainulainen K, Child A, Peltonen L, Steinmann B. 1994. Analyses of truncated fibrillin caused by a 366 bp deletion in the FBN1 gene resulting in Marfan syndrome. Biochem J 302(Pt 3):889-896.

Raghunath M, Kielty CM, Steinmann B. 1995. Truncated profibrillin of a Marfan patient is of apparent similar size as fibrillin: intracellular retention leads to over-N-glycosylation. J Mol Biol 248(5):901-909.

Ramirez F, Rifkin DB. 2009. Extracellular microfibrils: contextual platforms for TGFbeta and BMP signaling. Curr Opin Cell Biol 21(5):616-622.

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. J Comput Biol 4(3):311-323.

Reinhardt DP, Keene DR, Corson GM, Poschl E, Bachinger HP, Gambee JE, Sakai LY. 1996. Fibrillin-1: organization in microfibrils and structural properties. J Mol Biol 258(1):104-116.

Ritty TM, Broekelmann T, Tisdale C, Milewicz DM, Mecham RP. 1999. Processing of the fibrillin-1 carboxyl-terminal domain. J Biol Chem 274(13):8933-8940.

Robinson PN, Neumann LM, Demuth S, Enders H, Jung U, Konig R, Mitulla B, Muller D, Muschke P, Pfeiffer L and others. 2005. Shprintzen-Goldberg syndrome: fourteen new patients and a clinical analysis. Am J Med Genet A 135(3):251-262.

Robinson PN, Arteaga-Solis E, Baldock C, Collod-Beroud G, Booms P, De Paepe A, Dietz HC, Guo G, Handford PA, Judge DP, Kielty CM, Loeys B, Milewicz DM, Ney A, Ramirez F, Reinhardt DP, Tiedemann K, Whiteman P, Godfrey M. 2006. The molecular genetics of Marfan syndrome and related disorders. J Med Genet 43(10):769-787.

Roca X, Sachidanandam R, Krainer AR. 2003. Intrinsic differences between authentic and cryptic 5' splice sites. Nucleic Acids Res 31(21):6321-6333.

Royce PM and Steinmann UB. 2002. Connective tissue and its heritable disorders. John Wiley and Sons.

Sakai LY, Keene DR, Engvall E. 1986. Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. J Cell Biol 103(6 Pt 1):2499-2509.

Sakai LY, Keene DR, Glanville RW, Bachinger HP. 1991. Purification and partial characterization of fibrillin, a cysteine-rich structural component of connective tissue microfibrils. J Biol Chem 266(22):14763-14770.

Schimpf S, Fuhrmann N, Schaich S, Wissinger B. 2008. Comprehensive cDNA study and quantitative transcript analysis of mutant OPA1 transcripts containing premature termination codons. Hum Mutat 29(1):106-112.

Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 18(20):6097-6100.

Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. 2005. The FoldX web server: an online force field. Nucleic Acids Res 33(Web Server issue):W382-388.

Singh KK, Elligsen D, Liersch R, Schubert S, Pabst B, Arslan-Kirchner M, Schmidtke J. 2006a. Multi-exon out of frame deletion of the FBN1 gene leading to a severe juvenile onset cardiovascular phenotype in Marfan syndrome. J Mol Cell Cardiol 42(2):352-356.

Singh KK, Rommel K, Mishra A, Karck M, Haverich A, Schmidtke J, Arslan-Kirchner M. 2006b. TGFBR1 and TGFBR2 mutations in patients with features of Marfan syndrome and Loeys-Dietz syndrome. Hum Mutat 27(8):770-777.

Singh KK, Shukla PC, Schmidtke J. 2008. Conservation of 5'-upstream region of the FBN1 gene in primates. Eur J Hum Genet 16(7):869-872.

Skovby F, Gaustadnes M, Mudd SH. 2010. A revisit to the natural history of homocystinuria due to cystathionine beta-synthase deficiency. Mol Genet Metab 99(1):1-3.

Snead MP, Yates JR. 1999. Clinical and Molecular genetics of Stickler syndrome. J Med Genet 36(5):353-359.

Solis AS, Shariat N, Patton JG. 2008. Splicing fidelity, enhancers, and disease. Front Biosci 13:1926-1942.

Stalder L, Mühlemann O. 2008. The meaning of nonsense. Trends Cell Biol 18(7):315-321.

Streeten EA, Murphy EA, Pyeritz RE. 1987. Pulmonary function in the Marfan syndrome. Chest 91(3):408-412.

Summers KM, West JA, Peterson MM, Stark D, McGill JJ, West MJ. 2006. Challenges in the diagnosis of Marfan syndrome. Med J Aust 184(12):627-631.

Summers KM, Bokil NJ, Baisden JM, West MJ, Sweet MJ, Raggatt LJ, Hume DA. 2009. Experimental and bioinformatic characterisation of the promoter region of the Marfan syndrome gene, FBN1. Genomics 94(4):233-240.

ten Dijke P, Arthur HM. 2007. Extracellular control of TGFbeta signalling in vascular development and disease. Nat Rev Mol Cell Biol 8(11):857-869.

Treasure T. 2000. Cardivascular surgery for Marfan syndrome. Heart 84(6):674-678.

Van Camp G, Snoeckx RL, Hilgert N, van den Ende J, Fukuoka H, Wagatsuma M, Suzuki H, Smets RM, Vanhoenacker F, Declau F, Van de Heyning P, Usami S. 2006. A new autosomal recessive form of Stickler syndrome is caused by a mutation in the COL9A1 gene. Am J Hum Genet 79(3):449-457.

Vanita V, Singh JR, Singh D, Varon R, Robinson PN, Sperling K. 2007. A recurrent FBN1 mutation in an autosomal dominant ectopia lentis family of Indian origin. Mol Vis 13:2035-2040.

Vaughan CJ, Casey M, He J, Veugelers M, Henderson K, Guo D, Campagna R, Roman MJ, Milewicz DM, Devereux RB, Basson CT. 2001. Identification of a chromosome 11q23.2-q24 locus for familial aortic aneurysm disease, a genetically heterogeneous disorder. Circulation 103(20):2469-2475.

Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA 14(5):802-813.

Wang L, Guo DC, Cao J, Gong L, Kamm KE, Regalado E, Li L, Shete S, He WQ, Zhu MS, Offermanns S, Gilchrist D, Elefteriades J, Stull JT, Milewicz DM. 2010. Mutations in myosin light chain kinase cause familial aortic dissections. Am J Hum Genet 87(5):701-707.

Wilkinson MF. 2005. A new function for nonsense-mediated mRNA-decay factors. Trends Genet 21(3):143-148.

Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. 2002. Allelic variation in human gene expression. Science 297(5584):1143.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 11(2-3):377-394.

Yuan X, Downing AK, Knott V, Handford PA. 1997. Solution structure of the transforming growth factor beta-binding protein-like module, a domain associated with matrix fibrils. Embo J 16(22):6659-6666.

Zetoune AB, Fontanière S, Magnin D, Anczuków O, Buisson M, Zhang CX, Mazoyer S. 2008. Comparison of nonsense-mediated mRNA decay efficiency in various murine tissues. BMC Genet 9:83.

Zhang XH, Chasin LA. 2004. Computational definition of sequence motifs governing constitutive exon splicing. Genes Dev 18(11):1241-1250.

Zhang XH, Kangsamaksin T, Chao MS, Banerjee JK, Chasin LA. 2005. Exon inclusion is dependent on predictable exonic splicing enhancers. Mol Cell Biol 25(16):7323-7332.

Zhang Z, Lotti F, Dittmar K, Younis I, Wan L, Kasim M, Dreyfuss G. 2008. SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. Cell 133(4):572-574.

Zhu L, Vranckx R, Khau Van Kien P, Lalande A, Boisset N, Mathieu F, Wegman M, Glancy L, Gasc JM, Brunotte F, Bruneval P, Wolf JE, Michel JB, Jeunemaitre X. 2006. Mutations in myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. Nat Genet 38(3):343-349.

Acknowledgements

It is a pleasure to thank Prof. Wolfgang Berger who gave me the opportunity to do my thesis in his lab and as my doctoral father for discussions, useful comments, and the supervision of my work.

I would like to thank my supervisor Dr. Gábor Mátyás for this interesting dissertation project and the previous and current members of the Marfan-group for support.

Special thanks to Barbara Kloeckener-Gruissem, Christina Zeitz, and Ursula Forster. It was a very nice time sharing the office and workplace, discussing scientific and less scientific ideas, but also spending time together for fun and pleasure.

I would like to thank John Neidhardt, Nikolaus Schäfer, Lucas Mohn, Jurian Zürcher, Fabian Schmid, Sandra Brunner, Stephan Labs, Gaby Tanner, Ulrich Luhmann, Katharina Wycisk, Silke Feil, Mariana Wittmer, Jaya Balakrishnan, Esther Glaus, Philippe Reuge, Walther Hänseler, Ute Böttinger, Germaine Korner, Lea Sollfrank, Martin Fritzsche, Jadwiga Oczos, Britta Seebauer, Romain Da Costa, and all the members past and present I worked with during the previous years in the Berger laboratory. We had nice times discussing scientific problems, drinking coffee, beer, and wine, organizing barbeque evenings, football happenings, and trips. Thank you very much guys for the lot of fun and pleasure.

I would like to address my particular thanks to Beatrice Güdel and Monika Jenny for administrative and organizing help.

I also acknowledge Prof. Beat Steinmann, Angelika Schwarze, and the members of the Division of Metabolism and Molecular Pediatrics of the University Children's Hospital for cell culture and support.

I am also indebted to Dr. Deborah Bartholdi for taking of blood samples in every time I needed control samples.

I would like to thank my committee members, Prof. Thierry Hennet and Prof. Dieter Zimmermann for their support and helpful suggestions.

Many friends have helped me stay sane through these difficult years. Their support and care helped me overcome setbacks and stay focused on my graduate study. I greatly value their friendship and I deeply appreciate their belief in me.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of this dissertation.

Appendix 1

American Journal of Medical Genetics Part A 143A:1150–1158 (2007)

A Non-Ancestral *RPGR* Missense Mutation in Families With Either Recessive or Semi-Dominant X-Linked Retinitis Pigmentosa

Eyal Banin,[1] Liliana Mizrahi-Meissonnier,[1] Ruhama Neis,[1] Shira Silverstein,[2] István Magyar,[3] Dvorah Abeliovich,[2] Ronald Roepman,[4] Wolfgang Berger,[3] Thomas Rosenberg,[5] and Dror Sharon[1]*

[1]Department of Ophthalmology, Hadassah-Hebrew University Medical Center, Jerusalem, Israel
[2]Department of Human Genetics, Hadassah-Hebrew University Medical Center, Jerusalem, Israel
[3]Division of Medical Molecular Genetics and Gene Diagnostics, Institute of Medical Genetics, University of Zurich, Zurich, Switzerland
[4]Department of Human Genetics, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands
[5]Gordon Norrie Centre for Genetic Eye Diseases, National Eye Clinic for the Visually Impaired, Hellerup, Denmark

Received 28 March 2006; Accepted 24 November 2006

Most X-linked diseases show a recessive pattern of inheritance in which female carriers are unaffected. In X-linked retinitis pigmentosa (XLRP), however, both recessive and semi-dominant inheritance patterns have been reported. We identified an Israeli family with semi-dominant XLRP due to a missense mutation (p.G275S) in the *RPGR* gene. The mutation was previously reported in two Danish families with recessive XLRP. Obligate carriers from the two Danish families had no visual complaints and normal to slightly reduced retinal function, while those from the Israeli family suffered from high myopia, low visual acuity, constricted visual fields, and severely reduced electroretinogram (ERG) amplitudes. The disease-related *RPGR* haplotype of the Israeli family was found to be different from the one found in the two Danish families, indicating that the mutation arose twice independently on different X-chromosome backgrounds. A series of genetic analyses excluded skewed X-inactivation pattern, chromosomal abnormalities, distorted *RPGR* expression level, and mutations in candidate genes as the cause for the differences in disease severity of female carriers. To the best of our knowledge, this is the first detailed analysis of an identical mutation causing either a recessive or a semi-dominant X-linked pattern of disease in different families. Our results indicate that an additional gene (or genes), linked to *RPGR*, modulate disease expression in severely affected carriers. These may be related to the high myopia concomitantly found in affected carriers from the Israeli family. © 2007 Wiley-Liss, Inc.

Key words: electroretinography; eye disease; human genetics; mutation analysis; retinitis pigmentosa; X-linked disease

How to cite this article: Banin E, Mizrahi-Meissonnier L, Neis R, Silverstein S, Magyar I, Abeliovich D, Roepman R, Berger W, Rosenberg T, Sharon D. 2007. A non-ancestral *RPGR* missense mutation in families with either recessive or semi-dominant X-linked retinitis pigmentosa. Am J Med Genet Part A 143A:1150–1158.

INTRODUCTION

The vast majority of families with X-linked retinitis pigmentosa (XLRP) show a recessive mode of inheritance in which affected males usually manifest severe and rapid progressive visual loss with relatively early onset, while female carriers are asymptomatic or only minimally affected [Bird, 1975; Fishman, 1978]. Similar to other X-linked diseases, the variable and mild phenotype in carriers is mainly due to the random X-inactivation mechanism, in which one of the two X-chromosomes in female cells is usually randomly inactivated (the Lyonization effect [Lyon, 1961]). The fact that carriers of XLRP are usually asymptomatic implies that the expression of the normal allele in about half of the cells is sufficient for normal function of the affected tissue. However, a skewed or non-random inactivation in which the majority of cells express the mutant allele may lead to manifestation of disease [Brown and Robinson, 2000].

*Correspondence to: Dror Sharon, Department of Ophthalmology, Hadassah-Hebrew University Medical Center, Jerusalem, Israel.
E-mail: drorsharon@md.huji.ac.il
DOI 10.1002/ajmg.a.31642

American Journal of Medical Genetics Part A: DOI 10.1002/ajmg.a

Although most XLRP carriers are asymptomatic, careful fundoscopic and visual function testing often reveals mild retinal abnormalities as well as asymmetry between the eyes [Berson et al., 1979; Jacobson et al., 1989; Vajaranant et al., 2002]. In contrast to this usually mild, "recessive" pattern of disease expression in carriers, a few families with a semi-dominant pattern of XLRP have been reported in the literature [McGuire et al., 1995; Souied et al., 1997]. In these families, all or most carriers suffer from a widespread, clinically manifest retinal degeneration, which is less severe and of later onset as compared to affected males within the family, but much more pronounced than that usually seen in XLRP carriers. The mechanisms underlying the severe expression in females from these families are yet unknown.

Retinitis pigmentosa GTPase regulator (*RPGR*) is the major XLRP gene, responsible for 55–72% of cases [Vervoort et al., 2000; Breuer et al., 2002; Bader et al., 2003; Sharon et al., 2003]. The constitutive *RPGR* transcript is expressed in a wide variety of tissues and contains 19 exons [Meindl et al., 1996; Roepman et al., 1996]. Additional isoforms, produced by extensive alternative splicing, have been identified [Kirschner et al., 1999; Hong and Li 2002] including the RPGR-ORF15 variant which is a retina-enriched transcript that lacks exons 16–19 and contains ORF15 as its terminal exon [Vervoort et al., 2000]. The severity of retinal disease in affected males correlates with mutation location (ORF15 versus other exons) and mutation type within ORF15 [Sharon et al., 2003]. The function of the encoded RPGR protein is not yet known although the N-terminal region shows similarity to RCC1, a guanine nucleotide exchange factor for Ran GTPase [Meindl et al., 1996; Roepman et al., 1996]. RPGR is concentrated in photoreceptor connecting cilia and interacts with another ciliary protein, RPGRIP1 [Roepman et al., 2000; Hong et al., 2001]. While mutations causing recessive XLRP have been reported in different *RPGR* exons, those responsible for the semi-dominant form were found only in ORF15 [Rozet et al., 2002].

Aiming to better characterize the causes of semi-dominant XLRP inheritance, we performed a clinical and genetic analysis of families with either recessive or semi-dominant XLRP inheritance who share the same *RPGR* mutation. Interestingly, this mutation is not located within ORF15.

METHODS

Patient Recruitment

This study, which involved human patients, conformed to the tenets of the Declaration of Helsinki. Informed consent was obtained from all subjects after explanation of the nature and possible consequences of the study. The research was approved by the institutional review boards. DNA was purified from blood samples using the salting-out technique [Miller et al., 1988].

Clinical Evaluation

A full ophthalmologic examination including assessment of visual acuity, ocular motility, pupillary reaction, biomicroscopic slit lamp and dilated fundus examination was performed in all patients. Subsequently, kinetic perimetry (Goldman visual fields, targets V_{4e} and III_{4e}), color vision testing (Ishihara 38-plate and the Farnsworth-Munsell D-15 tests) and full-field electroretinography (ERG) were performed as previously described [Banin et al., 2003]. Electrooculography (EOG) was performed according to the ISCEV standard using bilateral skin electrodes on both canthi; the Arden ratio (light peak to dark trough) was derived [Marmor and Zrenner, 1993].

Screening for Mutations

Direct sequencing of PCR products was used to screen *RPGR* promoter region, exons 1–15, 15a, and ORF15 including the immediately flanking intron sequences for sequence changes as previously described [Sharon et al., 2003].

RT-PCR Analysis

Blood samples were collected into PAXgene tubes, incubated overnight at room temperature, and RNA was isolated using the PAXgene Blood RNA kit (PreAnalytix, Switzerland). Reverse Transcription was carried out using the Reverse-iT 1st Strand Synthesis kit (ABgene, UK). The ratio of allele-specific transcripts was quantified according to a procedure outlined by [Qiu et al., 2003] using primers specific to exons 7, 8, and 9, respectively (details are available upon request).

Haplotype Analysis

Thirteen microsatellite markers covering the short arm of the X-chromosome and a single nucleotide polymorphism (SNP) located within the *RPGR* gene were used for haplotype analysis.

X-Inactivation Pattern

DNA samples were obtained from white blood cells (WBC). The androgen-receptor (AR) gene X-inactivation assay using HpaII-sensitive methylation was performed as previously reported [Allen et al., 1992]. DNA samples from males were used as control for complete digestion. PCR products were run on an ABI Sequencer and analyzed using the GeneScan and Genotyper softwares (PE Biosystems, UK). As accepted in the literature, an inactivation level of at least 90% was considered significant [Beever et al., 2003].

American Journal of Medical Genetics Part A: DOI 10.1002/ajmg.a

RESULTS

Mutation Analysis of Family MOL0052

An Israeli family from Jewish Iranian origin (MOL0052), with a family history suggestive of XLRP, was identified. Obligatory female carriers manifested high myopia accompanied by relatively severe retinal and visual function abnormalities. The fundoscopic and functional changes were beyond those usually seen in carriers of XLRP, leading to the diagnosis of a semi-dominant pattern of inheritance. A screen for *RPGR* mutations revealed a missense substitution (c.g823a; p.Gly275Ser) in exon 8 which cosegregated with the retinal phenotype (Fig. 1). Interestingly, c.g823a has been previously reported as the cause of XLRP in two Danish patients, #2603 (family RP300304) and #2555 (family RP300318) [Roepman et al., 1996]. Despite extensive genealogical studies, we could not demonstrate any relationship between these two families. Our previous mutation analysis of these two patients was performed before the identification of ORF15. Sequencing analysis of this exon now revealed an inframe deletion (1307-1318del12) in *cis* with c.g823a. This inframe deletion, as well as many other ORF15 inframe deletions or insertions, was reported previously as a non-pathogenic polymorphism [Vervoort et al., 2000; Sharon et al., 2003].

Clinical Evaluation of XLRP Patients

We evaluated three affected males and six affected heterozygote females from MOL0052. On Funduscopy, affected males showed the classic signs of RP (Fig. 2A and B) including pallor of the optic discs, attenuation of vessels, and bone-spicule pigmentation (BSP). On fluorescein angiograms (FA), abnormal hyperfluorescence as well as areas of blocked fluorescence are seen (Fig. 2C and D). The youngest male patient (III-8) still had a visual acuity of 6/15–6/18, sizeable visual fields and measurable (yet markedly reduced) ERG responses (Table I). The two older males (III-3 and III-6) had very low visual acuities in one eye or both, tubular visual fields, and only the cone flicker responses were measurable. These were severely reduced and markedly delayed. In five out of six female patients, high congenital myopia and astigmatism (ranging from −8.5 D to −20 D) was present while in the affected males the refractive errors were less severe (Table I). Female patients also had different fundus findings, characterized by chorioretinal atrophy and peripheral pigment spots (Fig. 2E) that differed from the BSPs seen in the males. Visual dysfunction in the affected females, in general, was less severe than in the males. However, visual acuity impairment, loss of visual field, and markedly reduced and delayed ERG amplitudes were observed.

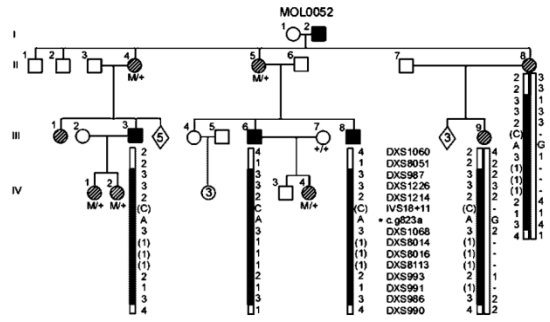

FIG. 1. Pedigree of family MOL0052 with semi-dominant XLRP inheritance. The c.g823a mutation (marked by asterisk) cosegregates with disease: Affected males were hemizygotes, affected females were heterozygotes, and unaffected family members carried only the wild-type allele. Haplotypes are depicted below the individual's symbol; black bar indicates the cosegregating disease haplotype surrounding the *RPGR* gene. Black fill: affected males; Diagonal lines: affected carrier females. The alleles in parenthesis were drawn from family members. The haplotype consisted of 13 microsatellite markers, one *RPGR* SNP (IVS18+11) and the pathogenic *RPGR* mutation (c.g823a). M: mutant *RPGR* allele; +: wild type allele.

Affected males in RP00304 and RP00318 presented typical features of XLRP with relatively severe disease (Table I). The refractive error varied from emmetropic to −6 D, with low astigmatism. One male, however, had high myopia of −8 D and −9 D. None of the female carriers in the Danish families had any visual complaints and displayed normal visual fields, and normal to only slightly reduced ERG amplitudes (Table I). On Funduscopy, the carrier females from family RP00304 had normal vessels and intact pigmentation, but did present typical tapetal reflexes in the posterior pole (Fig. 2F). A carrier female (IV-1) in family RP00318 also showed normal retinal vessels but the lower retinal periphery of both eyes revealed pigment irregularities and patches. The refractive errors of three carriers varied between emmetropia to −0.75 D. Carrier RP300304 IV-4, however, had hypermetropia and astigmatism in both eyes (Table I).

Genetic Analysis in Family MOL0052

The observation that a single *RPGR* mutation, p.G275S, is found in families with recessive XLRP as well as in a family with semi-dominant XLRP, prompted us to compare the disease-associated X-chromosome haplotype of the three families. A haplotype analysis of MOL0052 revealed only two putative recombination events leading to a large region (~43 cM) that cosegregates with the disease (Fig. 1). We also performed a haplotype analysis covering the whole X-chromosome in two affected females (II-8 who is highly myopic and her daughter, III-9, with no myopia) and did not identify any recombination events. Affected males from the two Danish families shared an identical haplotype in the vicinity of *RPGR*, which is different from the corresponding haplotype in MOL0052, indicating

American Journal of Medical Genetics Part A: DOI 10.1002/ajmg.a

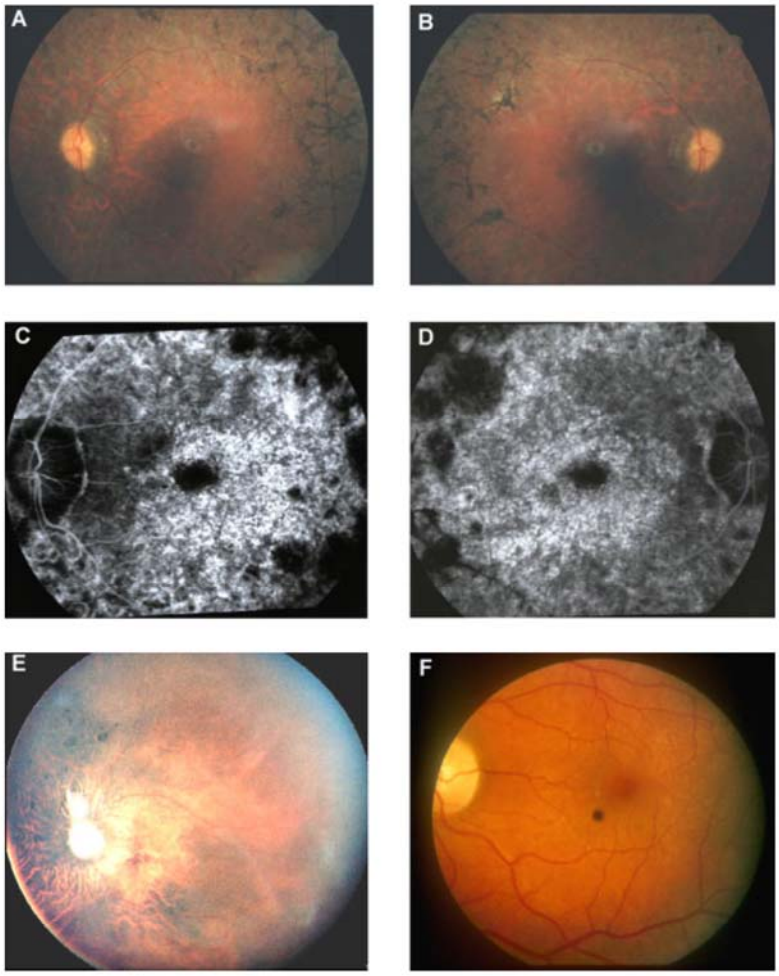

FIG. 2. Fundus photos (**A** and **B**) and fluorescein angiograms (**C** and **D**) of a male patient (MOL0052 III-6) demonstrating characteristic signs of RP including attenuated blood vessels, mild to moderate pallor of the optic discs, bone spicule-like pigment migration, hyperfluoresence in the posterior poles compatible with loss of RPE, and patches of blocked fluorescence temporally in the areas of pigment migration. Fundus photo (**E**) of an affected female carrier (MOL0052 II-8 at age 51) shows severe myopic changes, widespread areas of chorioretinal atrophy, and pigment clumps. Fundus photo (**F**) of a mildly affected Danish carrier (age 41) from family RP300304 shows the classic tapetal reflex.

that the mutation arose twice independently along human evolution and occurred on different chromosomal backgrounds. This result suggests that the p.G275S mutation is not the sole cause of disease in affected female carriers of family MOL0052, and another genetic factor, within *RPGR* or in another gene linked to *RPGR*, is involved in the pathogenesis of disease in female carriers. We therefore conducted a series of analyses aiming to decipher the mechanism which governs or modulates disease expression in affected female carriers.

First, since skewed (non-random) inactivation of the X-chromosome, with over-expression of the mutated gene in affected females, could theoretically be the cause, we used blood samples of affected males and females from the three studied families aiming to evaluate the inactivation pattern of the X-chromosome and the expression levels of the mutated *RPGR* allele versus the normal allele (Table II). We found no deviation from a random inactivation pattern (>90%) in eight out of the nine informative females (Table II). The only exception is patient II-4 (MOL0052) in which 98% inactivation of the normal allele was observed. The expression levels of the mutated *RPGR* allele were in agreement with the inactivation analysis (correlation of 0.97; $P = 3 \times 10^{-4}$). One carrier (RP300304 II-2), who was non-informative in the inactivation analysis, was found to express only the normal allele. Second, we used the Genomatix Promoter Inspection program (http://www.genomatix.de/) to identify the *RPGR* promoter region. The predicted promoter sequence upstream to exon 1 was then screened for mutations in one affected male (MOL0052 III-6) and no

American Journal of Medical Genetics Part A: DOI 10.1002/ajmg.a

TABLE I. Clinical Data of Affected Males and Females

Pt. #	Age	Sex	Refraction[a]	Visual acuity[b]	Color vision[c]	Goldman perimetry[d]	EOG[e] arden ratio	FFERG rod (μV)[f]	FFERG mixed cone-rod (μV)[g]	FFERG cone[h]	
										Amp (μV)	IT (mS)
MOL0052											
III-8	27	M	R emmetropic L emmetropic	R 6/18 L 6/15	R T L T	BE large MP scotomas	R 120% L 125%	R 0 L 0	R a: 45 b: 45 L a: 31 b: 43	R 9 L 26	42 42
III-6	34	M	R −5.00/−3.00 × 45° L −5.00/−3.00 × 20°	R 6/15 L 6/60	R T L T	BE Tubular (5–7°)	ND	R 0 L 0	R 0 L 0	R 23 L 14	42 41
III-3	37	M	R −6.00/−2.75 × 10° L −8.50/−0.50 × 62°	R 6/30 L 6/30	R T L T	R Tubular (5°) L immeasurable	R 138% L 100%	R 0 L 0	R 0 L 0	R 15 L 15	40 41
IV-4	3	F	R −9.00/−4.50 × 45° L −8.50/−5.00 × 125°	R FFM L FFM	ND	ND	ND	ND	BE Reduced (short protocol)	R 42 L 43	39 41
IV-2	6	F	R −9.50/−3.25 × 140° L −8.50 sphere	R 6/45 L 6/12	R normal L normal	ND	ND	R 105 L 240	R a: 17 b: 194 L a: 92 b: 308	R 48 L 51	37 37
III-9	20	F	R +1.75/−3.25 × 10° L +0.25/−3.00 × 175°	R 6/12 L 6/10	R normal L normal	ND	R 145% L 165%	R 92 L 82	R a: 68 b: 106 L a: 89 b: 157	R 20 L 14	36 36
II-8	49	F	R −10.00/−2.50 × 35° L −12.00/−1.25 × 135°	R 6/18 L 6/15	R normal L normal	BE large MP scotomas	R 150% L 163%	R 85 L 57	R a: 75 b: 134 L a: 57 b: 187	R 24 L 19	39 37
II-5	56	F	R −14.00/−2.50 × 30° L −17.00 sphere	R 6/18 L 6/24	R T L T	Constricted (30–40°)	R 165% L 147%	R 151 L 155	R a: 79 b: 177 L a: 87 b: 150	R 30 L 20	38 40
II-4	83	F	R −20.00 L −20.00	R LP L FC .3m	R ND L T	ND	ND	R 0 L 0	R 0 L 0	R 0 L 9	— 41
RP300304											
V-2	6	M	R −1.00/−1.25 × 24° L −1.00/−1.25 × 135°	R 6/12 L 6/12	ND	Normal	ND	ND	ND	ND	
IV-5	18	M	R −2.75/−1.25 × 35° L −2.50/−1.50 × 160°	R 6/9	R T L T	BE Tubular	ND	L 0	L 0	L 13	42
IV-1	36	M	R emmetropic L −0.50	R 6/24 L 6/18	R T L T	Tubular	ND	R 0	R 0	R 0	
IV-4	11	F	R +3.25/−3.50 × 27° L +5.75/−4.00 × 167°	R 6/9 L 6/9	ND	ND	ND	ND	ND	ND	
III-1	19	F	R −0.50/−0.50 × 100° L emmetropic	R 6/12 L 6/9	ND	Normal	ND	R 343	R blink artefact	R 56	33
RP300318											
V-2	17	M	R −8.00/−1.50 L −9.00/−1.50	R 6/18 L 6/24	ND	BE Moderate peripheral constriction	ND	ND	ND	ND	
V-5	34	M	R −2.00 L −1.50	R 6/18 L 6/18		BE Tubular with peripheral islands	ND	ND	ND	ND	
VI-3	9	M	R −5.00/−2.00 × 10° L −4.75/−1.75 × 0°	R 6/18 L 6/12	ND	Normal	ND	ND	ND	ND	
IV-1	51	F	R emmetropic L 0.00/−0.50 × 155°	R 6/6 L 6/6	R T L T	Normal	ND	R 212	R b: 287	R 49	30

R: right eye; L: left eye; BE: both eyes; ND: not done. For European patients only one eye was tested.

[a]Refraction: Sphere and cylinder (astigmatism) correction.

[b]Visual acuity (best corrected)—FFM: fix. follow, maintain; LP: light perception, no projection. FC: finger counting.

[c]Color vision—T: Tritanopia on Farnsworth Dichotomous Test panel D-15; ND: not done.

[d]Goldman perimetry: results for V4e target are described. MP: mid-peripheral.

[e]EOG: electrooculogram; normal arden ratio > 185%.

[f]FFERG: full field electroretinogram; 0: extinct. Rod response b-wave amplitude; normal for MOL0052 > 200μV, for RP300304 and RP300318 > 125μV.

[g]Mixed cone-rod response amplitudes; normal a-wave >90 μV; normal b-wave for MOL0052 >400 μV, for RP300304 and RP300318 >312 μV.

[h]30 Hz cone flicker response; normal amplitude (amp) for MOL0052 >60 μV, normal implicit time (IT) ≤ 33 msec. For RP300304 and RP300318: >74 μV, ≤30 msec.

American Journal of Medical Genetics Part A: DOI 10.1002/ajmg.a

TABLE II. Inactivation Pattern of the X-Chromosome and *RPGR* Expression Levels

Family	Pt. #	Age	Sex	RPGR genotype	AR genotype	Inactivation-mutant allele	Inactivation-normal allele/s	Expression (%)[a] mutant allele	Expression (%)[a] normal allele
MOL0052	III-7	33	F	+/+	261/288		28%; 72%		
MOL0052	III-4	36	F	+/+	255/273		44%; 56%		
MOL0052	III-3	37	M	G275S	270	N.A.	N.A.		
MOL0052	IV-4	3	F	G275S/+	270/288	19%	81%	70 (65–75)	30 (25–35)
MOL0052	IV-2	6	F	G275S/+	270/270	N.A.	N.A.	59 (57–62)	41 (38–43)
MOL0052	IV-1	13	F	G275S/+	270/270	N.A.	N.A.	66 (63–70)	34 (30–37)
MOL0052	III-9	20	F	G275S/+	270/270	N.A.	N.A.	92 (85–99)	8 (1–15)
MOL0052	II-8	49	F	G275S/+	270/282	62%	38%	44 (41–48)	56 (52–60)
MOL0052	II-5	56	F	G275S/+	270/255	75%	25%	38 (37–41)	62 (59–63)
MOL0052	II-4	83	F	G275S/+	270/255	2%	98%	100 (-)	not detectable
RP300304	III-4	30	F	G275S/+	267/282	80%	20%	23 (21–24)	77 (76–79)
RP300304	II-2	59	F	G275S/+	282/282	N.A.	N.A.	not detectable	100 (-)
RP300318	V-4	37	F	G275S/+	273/276	64%	36%	30 (28–32)	70 (68–72)
RP300318	IV-1	69	F	G275S/+	275/282	18%	82%	75 (74–76)	25 (24–26)

All affected males were hemizygote for both the *RPGR* mutation p.G275S and the AR allele 270 (MOL0052) or 282 (Danish families). Note that a recombination between *RPGR* and the *AR* gene was detected in individual RP300318 V-4. The female carriers MOL0052 III-9, IV-1, IV-2, and RP300304 II-2 were homozygote for the *AR* marker and thus an inactivation analysis was not applicable (N.A.).

[a]Lower and upper confidence limits of observed relative allele frequencies are given in parentheses ($P = 0.05$).

sequence changes were identified. Third, we studied the karyotype of one affected male (MOL0052 III-6) and one affected carrier (MOL0052 II-5) and did not identify any abnormalities (data not shown). Finally, since all carriers in MOL0052 were affected, we hypothesized that another sequence change in an *RPGR*-linked gene might be responsible for the relatively severe carrier phenotype. Aiming to identify candidate genes in the *RPGR* vicinity, we tabulated all known genes located up to 5 million bases from *RPGR* and gathered all the relevant information regarding their expression profile (mainly by SAGE and EST data) and predicted protein function. We identified 49 genes in this region and selected three as possible candidates. The first gene, BCL-6-interacting co-repressor (*BCOR*), is located 1.7Mb centromeric to *RPGR*. Distinct classes of *BCOR* mutations cause either Oculofaciocardiodental syndrome or Lenz microphthalmia syndrome [Ng et al., 2004]. The second gene, *DYNLT3*, is located 0.5Mb telomeric to *RPGR* and encodes a protein which is a member of the dynein light chain complex [DiBella et al., 2001]. *DYNLT3* interacts with the C-terminal cytoplasmic tail of rhodopsin, one of the major genes causing RP [Dryja et al., 1991], and its expression level in the retina is higher than in any other tissue (http://www.ncbi.nlm.nih.gov/SAGE/). The third gene, *RP2*, is a well-studied XLRP gene [Schwahn et al., 1998]. No mutations were identified in the open reading frame of these genes. In the *BCOR* gene, we identified a known non-pathogenic SNP (p.Asp420Asp; rs5917933).

DISCUSSION

The majority of X-linked diseases show a recessive mode of inheritance in which only a small proportion of carrier females are affected, usually to a much milder degree than hemizygous males. This is likely to be the result of the random nature of the well-documented Lyonization effect [Lyon, 1961]. Rarely, a different pattern of disease expression is seen in which female carriers are severely affected. This has historically been termed "dominant", "semi-dominant", or "intermediate". The use of the terms "recessive" and "dominant" in X-linked diseases is controversial, and a recent study proposed to discontinue the use of these terms and to simply categorize them all as "X-linked" [Dobyns et al., 2004]. This, however, might limit our ability to describe families in which a consistent pattern of disease expression occurs in carrier females. This distinction is even more crucial when describing families with XLRP in which both types of inheritance patterns have been reported due to mutations in the same gene, and as the data presented here suggest, even due to the same mutation.

The clinical findings among members of family MOL0052 are similar to previous reports of semi-dominant XLRP families, in which the carrier females are rather severely affected though still to a lesser degree than affected males [McGuire et al., 1995; Souied et al., 1997]. They may bear some resemblance to "grade 3" XLRP carriers [Grover et al., 2000]. This is in marked contrast to the clinical findings in the Danish carriers of the same mutation who closely resemble "grade 1" carriers in whom a tapetal reflex is present and degree of retinal dysfunction and deterioration was reported to be minimal [Grover et al., 2000]. A striking feature of the disease in female carriers from MOL0052 is the presence of high congenital myopia (often with significant astigmatism), while in affected males the myopia was less severe. Only one affected carrier in MOL0052 was not myopic (III-9) and our haplotype analysis indicates

American Journal of Medical Genetics Part A: DOI 10.1002/ajmg.a

that she inherited an intact X chromosome from her affected and highly myopic mother. It is therefore likely that the high degree of myopia is inherited in MOL0052, with reduced penetrance. High myopia has been reported in some carriers of XLRP, but was not necessarily associated with severe visual impairment [Fishman et al., 1986]. A previous study reported that high myopia was found consistently in only one out of nine families with semi-dominant XLRP inheritance [Souied et al., 1997]. However, the association of high myopia, astigmatism and marked impairment of retinal function was also described in female carriers of XLRP in two Japanese families [Yokoyama et al., 2001]. High myopia (at least −6 D) in itself was previously shown to correlate with reductions in full-field ERG amplitudes, but does not significantly affect implicit time [Westall et al., 2001]. It is reasonable to assume that part of the loss of retinal function in female carriers from family MOL0052 is due to their high myopia. The degree of retinal injury in these carrier females, however, exceeds the degree predicted by the myopia alone. In addition, cone flicker implicit time is delayed in all affected females, while it is usually within the normal range in highly myopic eyes without RP [Westall et al., 2001]. Interestingly, the asymmetry in retinal function that is often observed in carriers of XLRP [Jacobson et al., 1989] was only rarely present in G275S carriers. The marked visual impairment in the carrier females from MOL0052, as well as the high myopia associated with the carrier state, prompted us to seek whether an additional factor(s) modulates disease severity apart from the *RPGR* p.G275S mutation. This mutation is located within the RCC1 region and has been previously reported to interfere with the interaction between RPGR and RPGRIP1 [Roepman et al., 2000].

Beyond the Lyonization effect, the factors which determine the extent of retinal involvement and visual impairment in female carriers of XLRP are as yet unknown. The present study offers a rare opportunity to address this question, in view of the finding of an identical mutation in families with very different levels of disease severity in carriers. Thus, a number of possible causes for such differences could now be examined, and some ruled out: (a) As previously shown in many studies, most recessive XLRP families as well as most semi-dominant XLRP families have mutations in the *RPGR* gene, thus excluding the possibility that this specific gene in itself determines carrier severity; (b) as shown in the current study, the same *RPGR* mutation can be found in families with either recessive or semi-dominant XLRP inheritance, thus excluding a mutation-dependant mechanism. This is also supported by a previous study in which eight different null *RPGR*-ORF15 mutations were identified in nine families with semi-dominant XLRP [Rozet et al., 2002]. In line with our observation, two of these mutations have been previously described in families with recessive XLRP, but no clinical or genetic comparisons between these families have been reported; (c) is an additional genetic factor involved in determining disease severity? Our haplotype analysis revealed different haplotypes in the vicinity of *RPGR* for the Danish families versus MOL0052. This can be explained by a non-ancestral mutation mechanism, in which two different genetic events led to the same mutation, but on different chromosomal backgrounds. It is thus reasonable to assume that the consistent difference in phenotype is accounted for by a yet unknown genetic factor, which is linked to *RPGR* and therefore co-segregates with the retinal phenotype in female carriers. Such a genetic factor (or factors) may be common in Iranian Jews, perhaps acting by causing consistent skewed X-chromosome inactivation in which the mutated allele is active in most retinal cells. Alternatively, this factor/s may be involved in cell–cell communication or apoptosis. Since it is not possible to study X-inactivation in vivo directly in the retina, we studied the inactivation level in WBC of members of the three families. Only one female (MOL0052 II-4 at age of 83 years) showed a statistically significant skewed inactivation pattern, in which the normal allele was preferentially inactivated (98%). This is in line with a previous analysis showing that the incidence of skewed X-inactivation increases with age [Sharp et al., 2000]. It should be kept in mind, however, that the fact that random inactivation was maintained in peripheral blood cells does not necessarily imply that it occurred in the retina as well. As previously reported, although a significant association of X-inactivation ratios between tissues is usually present, wide variations are apparent in some cases [Sharp et al., 2000]. We therefore cannot rule-out the possibility that skewed X-inactivation did occur in the retina of these carrier females but is not evident in the WBC in their peripheral blood. Regrettably, there is no direct, reliable way to assess this in vivo. It was previously suggested that the pattern of the tapeto-retinal reflex in female carriers of XLRP (following image enhancement) correlates with the retinal mosaic of inactivation [Cideciyan and Jacobson, 1994], but the severity of disease in the affected females in MOL0052 did not allow this type of analysis. Seeking additional genes in the vicinity of *RPGR* which might contribute to the disease process, we screened three candidate genes but did not identify any potential mutations. However, since the chromosomal region that cosegregates with the disease is large, other genes may be involved in the severe retinal disease in affected females.

In summary, we report here of an identical *RPGR* mutation in families with either recessive or semi-dominant XLRP. To the best of our knowledge, this is the first detailed report showing that a single mutation which causes X-linked disease can be

American Journal of Medical Genetics Part A: DOI 10.1002/ajmg.a

that she inherited an intact X chromosome from her affected and highly myopic mother. It is therefore likely that the high degree of myopia is inherited in MOL0052, with reduced penetrance. High myopia has been reported in some carriers of XLRP, but was not necessarily associated with severe visual impairment [Fishman et al., 1986]. A previous study reported that high myopia was found consistently in only one out of nine families with semi-dominant XLRP inheritance [Souied et al., 1997]. However, the association of high myopia, astigmatism and marked impairment of retinal function was also described in female carriers of XLRP in two Japanese families [Yokoyama et al., 2001]. High myopia (at least −6 D) in itself was previously shown to correlate with reductions in full-field ERG amplitudes, but does not significantly affect implicit time [Westall et al., 2001]. It is reasonable to assume that part of the loss of retinal function in female carriers from family MOL0052 is due to their high myopia. The degree of retinal injury in these carrier females, however, exceeds the degree predicted by the myopia alone. In addition, cone flicker implicit time is delayed in all affected females, while it is usually within the normal range in highly myopic eyes without RP [Westall et al., 2001]. Interestingly, the asymmetry in retinal function that is often observed in carriers of XLRP [Jacobson et al., 1989] was only rarely present in G275S carriers. The marked visual impairment in the carrier females from MOL0052, as well as the high myopia associated with the carrier state, prompted us to seek whether an additional factor(s) modulates disease severity apart from the *RPGR* p.G275S mutation. This mutation is located within the RCC1 region and has been previously reported to interfere with the interaction between RPGR and RPGRIP1 [Roepman et al., 2000].

Beyond the Lyonization effect, the factors which determine the extent of retinal involvement and visual impairment in female carriers of XLRP are as yet unknown. The present study offers a rare opportunity to address this question, in view of the finding of an identical mutation in families with very different levels of disease severity in carriers. Thus, a number of possible causes for such differences could now be examined, and some ruled out: (a) As previously shown in many studies, most recessive XLRP families as well as most semi-dominant XLRP families have mutations in the *RPGR* gene, thus excluding the possibility that this specific gene in itself determines carrier severity; (b) as shown in the current study, the same *RPGR* mutation can be found in families with either recessive or semi-dominant XLRP inheritance, thus excluding a mutation-dependant mechanism. This is also supported by a previous study in which eight different null *RPGR*-ORF15 mutations were identified in nine families with semi-dominant XLRP [Rozet et al., 2002]. In line with our observation, two of these mutations have been previously described in families with recessive XLRP, but no clinical or genetic comparisons between these families have been reported; (c) is an additional genetic factor involved in determining disease severity? Our haplotype analysis revealed different haplotypes in the vicinity of *RPGR* for the Danish families versus MOL0052. This can be explained by a non-ancestral mutation mechanism, in which two different genetic events led to the same mutation, but on different chromosomal backgrounds. It is thus reasonable to assume that the consistent difference in phenotype is accounted for by a yet unknown genetic factor, which is linked to *RPGR* and therefore co-segregates with the retinal phenotype in female carriers. Such a genetic factor (or factors) may be common in Iranian Jews, perhaps acting by causing consistent skewed X-chromosome inactivation in which the mutated allele is active in most retinal cells. Alternatively, this factor/s may be involved in cell–cell communication or apoptosis. Since it is not possible to study X-inactivation in vivo directly in the retina, we studied the inactivation level in WBC of members of the three families. Only one female (MOL0052 II-4 at age of 83 years) showed a statistically significant skewed inactivation pattern, in which the normal allele was preferentially inactivated (98%). This is in line with a previous analysis showing that the incidence of skewed X-inactivation increases with age [Sharp et al., 2000]. It should be kept in mind, however, that the fact that random inactivation was maintained in peripheral blood cells does not necessarily imply that it occurred in the retina as well. As previously reported, although a significant association of X-inactivation ratios between tissues is usually present, wide variations are apparent in some cases [Sharp et al., 2000]. We therefore cannot rule-out the possibility that skewed X-inactivation did occur in the retina of these carrier females but is not evident in the WBC in their peripheral blood. Regrettably, there is no direct, reliable way to assess this in vivo. It was previously suggested that the pattern of the tapeto-retinal reflex in female carriers of XLRP (following image enhancement) correlates with the retinal mosaic of inactivation [Cideciyan and Jacobson, 1994], but the severity of disease in the affected females in MOL0052 did not allow this type of analysis. Seeking additional genes in the vicinity of *RPGR* which might contribute to the disease process, we screened three candidate genes but did not identify any potential mutations. However, since the chromosomal region that cosegregates with the disease is large, other genes may be involved in the severe retinal disease in affected females.

In summary, we report here of an identical *RPGR* mutation in families with either recessive or semi-dominant XLRP. To the best of our knowledge, this is the first detailed report showing that a single mutation which causes X-linked disease can be

American Journal of Medical Genetics Part A: DOI 10.1002/ajmg.a

associated with very different levels of disease severity in female carriers. Based on our genetic analysis, we predict that an additional modifier gene (or genes), linked to *RPGR*, is responsible for determining severity of disease in affected female carriers.

ACKNOWLEDGMENTS

The authors thank Ms. Lina Bida and Mrs. Silke Feil for technical assistance, Dr. Alex Obolensky for performing the visual field tests, Mr. Erik Kann for genealogical assistance, and the patients for their cooperation. This study was supported by the Israeli Woman Health Foundation, the Yedidut Research Grant, and the Ulla and Bernt Hjejles Memorial Foundation.

REFERENCES

Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. 1992. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. Am J Hum Genet 51:1229–1239.

Bader I, Brandau O, Achatz H, Apfelstedt-Sylla E, Hergersberg M, Lorenz B, Wissinger B, Wittwer B, Rudolph G, Meindl A, Meitinger T. 2003. X-linked retinitis pigmentosa: RPGR mutations in most families with definite X linkage and clustering of mutations in a short sequence stretch of exon O RF15. Invest Ophthalmol Vis Sci 44:1458–1463.

Banin E, Shalev RS, Obolensky A, Neis R, Chowers I, Gross-Tsur V. 2003. Retinal function abnormalities in patients treated with vigabatrin. Arch Ophthalmol 121:811–816.

Beever CL, Penaherrera MS, Langlois S, Robinson WR. 2003. X chromosome inactivation patterns in Russell-Silver syndrome patients and their mothers. Am J Med Genet Part A 123A:231–235.

Berson EL, Rosen JB, Simonoff EA. 1979. Electroretinographic testing as an aid in detection of carriers of X-chromosome-linked retinitis pigmentosa. Am J Ophthalmol 87:460–468.

Bird AC. 1975. X-linked retinitis pigmentosa. Br J Ophthalmol 59:177–199.

Breuer DK, Yashar BM, Filippova E, Hiriyanna S, Lyons RH, Mears AJ, Asaye B, Acar C, Vervoort R, Wright AF, Musarella MA, Wheeler P, MacDonald I, Iannaccone A, Birch D, Hoffman DR, Fishman GA, Heckenlively JR, Jacobson SG, Seving PA, Swaroop A. 2002. A comprehensive mutation analysis of RP2 and RPGR in a north american cohort of families with X-linked retinitis pigmentosa. Am J Hum Genet 70:1545–1554.

Brown CJ, Robinson WP. 2000. The causes and consequences of random and non-random X chromosome inactivation in humans. Clin Genet 58:353–363.

Cideciyan AV, Jacobson SG. 1994. Image analysis of the tapetal-like reflex in carriers of X-linked retinitis pigmentosa. Invest Ophthalmol Vis Sci 35:3812–3824.

DiBella LM, Benashski SE, Tedford HW, Harrison A, Patel-King RS, King SM. 2001. The Tctex1/Tctex2 class of dynein light chains. Dimerization, differential expression, and interaction with the LC8 protein family. J Biol Chem 276:14366–14373.

Dobyns WB, Filauro A, Tomson BN, Chan AS, Ho AW, Ting NT, Oosterwijk JC, Ober C. 2004. Inheritance of most X-linked traits is not dominant or recessive, just X-linked. Am J Med Genet Part A 129A:136–143.

Dryja TP, Hahn LB, Cowley GS, McGee TL, Berson EL. 1991. Mutation spectrum of the rhodopsin gene among patients with autosomal dominant retinitis pigmentosa. Proc Natl Acad Sci USA 88:9370–9374.

Fishman GA. 1978. Retinitis pigmentosa. Genetic percentages. Arch Ophthalmol 96:822–826.

Fishman GA, Weinberg AB, McMahon TT. 1986. X-linked recessive retinitis pigmentosa. Clinical characteristics of carriers. Arch Ophthalmol 104:1329–1335.

Grover S, Fishman GA, Anderson RJ, Lindeman M. 2000. A longitudinal study of visual function in carriers of X-linked recessive retinitis pigmentosa. Ophthalmology 107:386–396.

Hong DH, Li T. 2002. Complex expression pattern of RPGR reveals a role for purine-rich exonic splicing enhancers. Invest Ophthalmol Vis Sci 43:3373–3382.

Hong DH, Yue G, Adamian M, Li T. 2001. Retinitis pigmentosa GTPase regulator (RPGRr)-interacting protein is stably associated with the photoreceptor ciliary axoneme and anchors RPGR to the connecting cilium. J Biol Chem 276:12091–12099.

Jacobson SG, Yagasaki K, Feuer WJ, Roman AJ. 1989. Interocular asymmetry of visual function in heterozygotes of X-linked retinitis pigmentosa. Exp Eye Res 48:679–691.

Kirschner R, Rosenberg T, Schultz-Heienbrok R, Lenzner S, Feil S, Roepman R, Cremers FP, Ropers HH, Berger W. 1999. RPGR transcription studies in mouse and human tissues reveal a retina-specific isoform that is disrupted in a patient with X-linked retinitis pigmentosa. Hum Mol Genet 8:1571–1578.

Lyon MF. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). Nature 190:372–373.

Marmor MF, Zrenner E. 1993. Standard for clinical electro-oculography. International Society for Clinical Electrophysiology of Vision. Arch Ophthalmol 111:601–604.

McGuire RE, Sullivan LS, Blanton SH, Church MW, Heckenlively JR, Daiger SP. 1995. X-linked dominant cone-rod degeneration: Linkage mapping of a new locus for retinitis pigmentosa (RP 15) to Xp22.13-p22.11. Am J Hum Genet 57:87–94.

Meindl A, Dry K, Herrmann K, Manson F, Ciccodicola A, Edgar A, Carvalho MR, Achatz H, Hellebrand H, Lennon A, Migliaccio C, Porter K, Zrenner E, Bird A, Jay M, Lorenz B, Wittwer B, D'Urso M, Meitinger T, Wright A. 1996. A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). Nat Genet 13:35–42.

Miller SA, Dykes DD, Polesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucl Acids Res 16:1215.

Ng D, Thakker N, Corcoran CM, Donnai D, Perveen R, Schneider A, Hadley DW, Tifft C, Zhang L, Wilkie AO, van der Smagt JJ, Gorlin RJ, Burgess SM, Bardwell VJ, Black GC, Biesecker LG. 2004. Oculofaciocardiodental and Lenz microphthalmia syndromes result from distinct classes of mutations in BCOR. Nat Genet 36:411–416.

Qiu P, Soder GJ, Sanfiorenzo VJ, Wang L, Greene JR, Fritz MA, Cai XY. 2003. Quantification of single nucleotide polymorphisms by automated DNA sequencing. Biochem Biophys Res Commun 309:331–338.

Roepman R, van Duijnhoven G, Rosenberg T, Pinckers AJ, Bleeker-Wagemakers LM, Bergen AA, Post J, Beck A, Reinhardt R, Ropers HH, Cremers FP, Berger W. 1996. Positional cloning of the gene for X-linked retinitis pigmentosa 3: Homology with the guanine-nucleotide-exchange factor R CC1. Hum Mol Genet 5:1035–1041.

Roepman R, Bernoud-Hubac N, Schick DE, Maugeri A, Berger W, Ropers HH, Cremers FP, Ferreira PA. 2000. The retinitis pigmentosa GTPase regulator (RPGR) interacts with novel transport-like proteins in the outer segments of rod photoreceptors. Hum Mol Genet 9:2095–2105.

Rozet JM, Perrault I, Gigarel N, Souied E, Ghazi I, Gerber S, Dufier JL, Munnich A, Kaplan J. 2002. Dominant X linked retinitis pigmentosa is frequently accounted for by truncating mutations in exon ORF15 of the RPGR gene. J Med Genet 39:284–285.

American Journal of Medical Genetics Part A: DOI 10.1002/ajmg.a

Schwahn U, Lenzner S, Dong J, Feil S, Hinzmann B, van Duijnhoven G, Kirschner R, Hemberger M, Bergen AA, Rosenberg T, Pinckers AJ, Fundele R, Rosenthal A, Cremers FP, Ropers HH, Berger W. 1998. Positional cloning of the gene for X-linked retinitis pigmentosa 2. Nat Genet 19:311–313.

Sharon D, Sandberg MA, Rabe VW, Stillberger M, Dryja TP, Berson EL. 2003. RP2 and RPGR mutations and clinical correlations in patients with X-linked retinitis pigmentosa. Am J Hum Genet 73:1131–1146.

Sharp A, Robinson D, Jacobs P. 2000. Age- and tissue-specific variation of X chromosome inactivation ratios in normal women. Hum Genet 107:343–349.

Souied E, Segues B, Ghazi I, Rozet JM, Chatelin S, Gerber S, Perrault I, Michel-Awad A, Briard ML, Plessis G, Dufier JL, Munnich A, Kaplan J. 1997. Severe manifestations in carrier females in X linked retinitis pigmentosa. J Med Genet 34:793–797.

Vajaranant TS, Seiple W, Szlyk JP, Fishman GA. 2002. Detection using the multifocal electroretinogram of mosaic retinal dysfunction in carriers of X-linked retinitis pigmentosa. Ophthalmology 109:560–568.

Vervoort R, Lennon A, Bird AC, Tulloch B, Axton R, Miano MG, Meindl A, Meitinger T, Ciccodicola A, Wright AF. 2000. Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. Nat Genet 25:462–466.

Westall CA, Dhaliwal HS, Panton CM, Sigesmun D, Levin AV, Nischal KK, Heon E. 2001. Values of electroretinogram responses according to axial length. Doc Ophthalmol 102: 115–130.

Yokoyama A, Maruiwa F, Hayakawa M, Kanai A, Vervoort R, Wright AF, Yamada K, Niikawa N, Naoi N. 2001. Three novel mutations of the RPGR gene exon ORF15 in three Japanese families with X-linked retinitis pigmentosa. Am J Med Genet 104:232–238.

ERRATUM

AMERICAN JOURNAL OF medical genetics PART A

Erratum[Q1] to "A Non-Ancestral RPGR Missense Mutation in Families With Either Recessive or Semi-Dominant X-Linked Retinitis Pigmentosa"

Dror Sharon*

Hadassah-Hebrew University Medical Center, Department of Ophthalmology, Jerusalem, Israel

Received 29 January 2009; Accepted 23 March 2009

American Journal of Medical Genetics Part A 143A:1150–1158 (2007).

Correction to the Acknowledgment Section. In the article titled A Non-Ancestral RPGR Missense Mutation in Families With Either Recessive or Semi-Dominant X-Linked Retinitis Pigmentosa, published in the June 2007 issue of AJMG part A (2007;143A:1150–1158), we regret to have omitted from the acknowledgment section the following contribution:

We are very grateful to Dr. Gábor Mátyás for experimental advice, constructive discussions, and helpful comments regarding quantitative sequencing.

Contributions of authors to the manuscript "A Non-Ancestral *RPGR* Missense Mutation in Families with Either Recessive or Semi-Dominant X-Linked Retinitis Pigmentosa"

Eyal Banin	Conceptual planning, design and supervision of the study, interpretation of data, writing and editing of the manuscript
Liliana Mizrahi-Meissonnier	Providing of technical help, editing of the manuscript
Ruhama Neis	Providing of technical help, editing of the manuscript
Shira Silverstein	Providing of technical help, editing of the manuscript
István Magyar	Quantitative transcript analysis, editing of the manuscript
Dvorah Abeliovich	Conceptual input, editing of the manuscript
Ronald Roepman	Conceptual input, editing of the manuscript
Wolfgang Berger	Conceptual input, interpretation of data and results, editing of the manuscript
Thomas Rosenberg	Conceptual input, sending patient`s material, editing of the manuscript
Dror Sharon	Conceptual planning, design and supervision of the study, interpretation of data, writing and editing of the manuscript

Appendix 2

Biochemistry and Molecular Biology

Alterations of the 5′Untranslated Region of *SLC16A12* Lead to Age-Related Cataract

Jurian Zuercher,[1] *John Neidhardt,*[1] *Istvan Magyar,*[1] *Stephan Labs,*[1] *Anthony T. Moore,*[2,3] *Felix C. Tanner,*[4] *Naushin Waseem,*[3] *Daniel F. Schorderet,*[5] *Francis L. Munier,*[6] *Shomi Bhattacharya,*[3] *Wolfgang Berger,*[1] *and Barbara Kloeckener-Gruissem*[1,7]

Purpose. Knowledge of genetic factors predisposing to age-related cataract is very limited. The aim of this study was to identify DNA sequences that either lead to or predispose for this disease.

Methods. The candidate gene *SLC16A12*, which encodes a solute carrier of the monocarboxylate transporter family, was sequenced in 484 patients with cataract (134 with juvenile cataract, 350 with age-related cataract) and 190 control subjects. Expression studies included luciferase reporter assay and RT-PCR experiments.

Results. One patient with age-related cataract showed a novel heterozygous mutation (c.-17A>G) in the 5′untranslated region (5′UTR). This mutation is in *cis* with the minor G-allele of the single nucleotide polymorphism (SNP) rs3740030 (c.-42T/G), also within the 5′UTR. Using a luciferase reporter assay system, a construct with the patient's haplotype caused a significant upregulation of luciferase activity. In comparison, the SNP G-allele alone promoted less activity, but that amount was still significantly higher than the amount of the common T-allele. Analysis of *SLC16A12* transcripts in surrogate tissue demonstrated striking allele-specific differences causing 5′UTR heterogeneity with respect to sequence and quantity. These differences in gene expression were mirrored in an allele-specific predisposition to age-related cataract, as determined in a Swiss population (odds ratio approximately 2.2; confidence intervals, 1.23–4.3).

Conclusions. The monocarboxylate transporter SLC16A12 may contribute to age-related cataract. Sequences within the 5′UTR modulate translational efficiency with pathogenic consequences. (*Invest Ophthalmol Vis Sci.* 2010;51:3354–3361) DOI:10.1167/iovs.10-5193

Cataract is the clouding of the eye's lens that impairs normal vision. It is estimated that cataract accounts for 17 million cases of blindness worldwide, with approximately half of all cases occurring in Asia and Africa.[1,2] Different criteria—age of onset, morphologic features, and mode of inheritance—can be used to classify the various forms of cataracts. Based on the age of onset, one distinguishes between childhood (congenital and juvenile) cataract and age-related cataract (ARC), but this criterion does not necessarily indicate etiology.[3] Genetic predisposition plays a crucial role in childhood cataract.[1,2] Congenital and juvenile forms of cataract show wide heterogeneity with respect to genetic and phenotypic aspects.[4] A number of mutations in approximately 20 genes have been described as causing childhood cataract.[3,5–7]

Approximately 80% of all cataracts are age-related and idiopathic. Depending on the location of the opacity within the lens, ARC is termed cortical, nuclear, or subcapsular. There are also forms of mixed cataract that feature more than one morphological sign. In general, maintenance of an intact, transparent lens requires balanced homeostasis of metabolic components.[8] ARC is considered a multifactorial disease in which environmental components and genetic predisposition contribute to the development of the pathologic condition. Interactions between these factors are likely, and knowledge of the cause of ARC may provide crucial information for the prevention of and potential therapy for the disease. Among environmental risk factors are smoking, exposure to UV-B radiation, and alcohol.[9] In addition, physiological conditions such as age, sex, diabetes, high body mass index, persistent intraocular inflammation, prolonged corticosteroid administration, and oxidative damage seem to promote the development of ARC.[9] In light of this complexity, knowledge of genetic risk factors is still scarce. Variants of the detoxifying enzymes arylamine *N*-acetyltransferase-2 and glutathione-*S*-transferase (GST) were found to be associated with ARC.[10–12] Furthermore, two sequence variants in the vicinity of the 3′ end of the gene for Eph-receptor tyrosine kinase type A2 (*EPHA2*) were shown to associate with both childhood and age-related forms of cataract.[13] Recently, a mutation in the gene encoding αA-crystallin was reported to be associated with ARC because of the loss of chaperone-like activity.[14] SNP-based allele frequencies and, consequently, association with a disease phenotype vary often among different ethnic groups, exemplified by a sequence variant in the gene encoding galactokinase, an enzyme involved in galactose metabolism[15,16] and GST.[10,11,17] Similarly, heat-shock transcription factor 4 may be involved in ARC in an Asian population.[18]

From the [1]Division of Medical Molecular Genetics and Gene Diagnostics, Institute of Medical Genetics, University of Zurich, Zurich, Switzerland; [2]Moorfields Eye Hospital London, London, United Kingdom; [3]UCL-Institute of Ophthalmology, London, United Kingdom; the [4]Department of Cardiology, Cardiovascular Center, University of Zurich, Zurich, Switzerland; [5]IRO-Institute for Research in Ophthalmology, EPFL-École polytechnique fédérale of Lausanne and University of Lausanne, Lausanne, Switzerland; [6]Jules Gonin Eye Hospital, Faculté de Biologie et Médecine de L'Université de Lausanne, Switzerland; and the [7]Department of Biology, ETH Zurich, Zurich, Switzerland.

Supported by the National Institute for Health Research Biomedical Research Centre for Ophthalmology and The Special Trustees of Moorfields Eye Hospital London (SB); Wellcome Trust and National Institute for Health Research (UK); Moorfields Eye Hospital BMRC (ATM); and Swiss National Science Foundation Grant 320030_127558 (FLM, DFS).

Submitted for publication January 11, 2010; revised February 4, 2010; accepted February 9, 2010.

Disclosure: **J. Zuercher**, None; **J. Neidhardt**, None; **I. Magyar**, None; **S. Labs**, None; **A.T. Moore**, None; **F.C. Tanner**, None; **N. Waseem**, None; **D.F. Schorderet**, None; **F.L. Munier**, None; **S. Bhattacharya**, None; **W. Berger**, None; **B. Kloeckener-Gruissem**, None

Corresponding author: Barbara Kloeckener-Gruissem, Division of Medical Molecular Genetics and Gene Diagnostics, Institute of Medical Genetics, University of Zurich, Schorenstrasse 16, 8603 Schwerzenbach, Switzerland; kloeckener@medgen.uzh.ch.

Investigative Ophthalmology & Visual Science, July 2010, Vol. 51, No. 7

To the list of genes involved in childhood cataract[19] we recently added *SLC16A12*, which encodes a monocarboxylate transporter.[6] Although the substrate of this transporter is not yet known, its importance for the establishment or maintenance of homeostasis was suggested because a premature termination codon in *SLC16A12* leads to juvenile cataract and renal glucosuria.[6] Based on this proposed function of the transporter, we speculated that insufficient activity could interfere with the maintenance of homeostatic conditions within the lens and could lead to ARC. Now we report the effects on ARC of two sequence alterations in the 5′untranslated region (5′UTR) of *SLC16A12*. The importance of 5′UTR sequences for translational regulation has been demonstrated in several other genes.[20–23] Approximately 10% of all coding genes are regulated at the mRNA level and, hence, influence translational efficiency. Disturbance of the regulation of the translational machinery leads to perturbed cellular metabolism and may tilt the physiological balance from healthy to diseased states, as occurs in breast cancer, Alzheimer's disease, bipolar affective disorder, fragile X-syndrome, and others.[20,22] Our work focuses on sequence alterations within the 5′UTR of *SLC16A12* and offers an explanation for the development of ARC.

Materials and Methods

Patients

Patients with childhood cataract and ARC, including cortical, nuclear, posterior subcapsular, and mixed types of cataract, were seeking ophthalmologic examination in Switzerland. Subjects from among the general population in Switzerland served as controls. Information on patient's condition of hypertension, diabetes mellitus, smoking behavior, alcohol consumption, and exposure to UV-B radiation was not available. Patients gave written informed consent for participation in scientific research. All experiments involving human subjects were conducted according to the principles expressed in the Declaration of Helsinki.

DNA Analysis

DNA was prepared either by the precipitation method (Gentra Kit; (Qiagen, Hilden, Germany) or by magnetic bead technology (Chemagen, Aachen, Germany). For PCR, approximately 50 ng template DNA and primers (Table 1) were cycled 35 times with annealing and extension temperatures of 60°C and 72°C, respectively, lasting 1 minute each. DNA sequencing was performed with commercially available technology (Applied Biosystems, Rotkreuz, Switzerland).

RNA Analysis

RNA from vascular smooth muscle cell (VSMC) cultures was isolated (All Prep DNA RNA Mini Kit; Qiagen, Hilden, Germany). RNA was evaluated with a bioanalyzer (2100 Bioanalyzer; Agilent Technologies, Palo Alto, CA). Two-step RT-PCR required cDNA synthesis (Superscript III; Invitrogen, Basel, Switzerland). Standard RT-PCR conditions were applied (Hotfire *Taq* Polymerase; Solis Biodyne, Tartu, Estonia) for 1 minute at an annealing temperature of 58°C, 1 minute elongation time at 72°C, and 39 cycles. PCR products were analyzed by 1% agarose gel electrophoresis. Quantitative sequencing of the *SLC16A12* c.-42T>G variant was performed as described.[24] Briefly, to determine the correction factors, genomic DNA (gDNA) was amplified in duplicate, and each amplicon was sequenced in eight different reactions. For RNA, one-step RT-PCR (Qiagen) was performed in triplicate, and each amplicon was sequenced eight times. Potential splice sites were sought with the online tool ESEfinder 3.0 (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home).[25,26]

In Silico Analysis of 5′UTR Variants on RNA Folding

Putative RNA folding structures were predicted using Mfold with standard settings (http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi).[27] RNA structures were predicted for *SLC16A12* (CL reverse NM_213606) and for its 5′UTR alone.

Plasmids, Vectors, and Cloning

Exon 3 of *SLC16A12* contains part of the 5′UTR and the initiation ATG followed by 36 encoded amino acids. A fragment of the 118-bp *SLC16A12* sequence (5′-GCCAGGTAGCGTTCTATGCCAACCTTGAATGCCATCAGGAAGTCACTGGACAGCAAACTCTTCCAAGATCATAACTT[**G/T**]GGCTGTTGGAGCAACCTGGAAAAG[**A/G**]AGAAAAAAGAAAAACC-3′ was cloned in front of the luciferase gene in pGL3 control vector (Promega, Madison, WI) with *Hin*dIII and *Nco*I restriction enzymes. For this purpose, exon 3 was PCR amplified using an upstream primer (Table 1) that contained a *Hin*dIII restriction site at its 5′ end and a downstream primer (Table 1). Genomic DNA heterozygous for the SNP rs37400030 and genomic DNA from the cataract patient with c.-42G/T and c. -17A/G served as templates. PCR amplicons and vector pGL3 were digested with restriction enzymes *Hin*dIII and *Nco*I and were ligated with DNA ligase (Promega). Constructs were verified by DNA sequence analysis.

Mammalian Cell Culture Experiments

Cells (10^4 HEK293T) were seeded on 96-well plates in 0.5 mL DMEM, 10% FBS, and 1% penicillin/streptomycin and were transfected 24 hours later. Plasmid DNA (615 ng [600 ng pFirefly construct, 15 ng pRenilla construct]) dissolved in 2 μL calcium chloride (2.5 M) and 10 μL HeBS (2×) was used for each transfection (20 μL transfection mix/well). Cells were harvested 48 hours after transfection, and Luciferase activities were measured with a Luciferase reporter assay system (Dual Glo; Promega Dual Glo) and a luminometer (Luminoskan Ascent; Thermo-Labsystems, Egelsbach, Germany).

Table 1. Primer Characteristics

Primer Name	Sequence (5′-3′)	Purpose
Intronic primers 3F	gtctgccccagtctagtattca	Genomic DNA sequencing
Intronic primers 3R	cggaaatacacacacaccaca	Genomic DNA sequencing
CL forward	ATGCaagcttGCCAGGTAGCGTTCTATGCC	Cloning
CL reverse	CGGAAATACACACACACCACA	Cloning
SLC16A12RT_ex1_2f	cggtggctcagATACAGGAT	RT-PCR
SLC16A12RT_ex4r	aaacagccagccacaatcat	RT-PCR
RTPRC_2F	GTGTGACCATGCTCTGTGct	RT-PCR
RTPRC_4R	AAGACAAAGCCCCCAAGAAT	RT-PCR
RTPCR_3F	caggaagtcactggacagca	RT-PCR
RTPCR_5R	gcatgattcccacttgacag	RT-PCR
cSLC16A12_1F	CCCTCTTCCCTCTCCCTGA	RT-PCR
cSLC16A12_3R	TGTGCAGATGGTAACAAGGAAAC	RT-PCR
SLC16A12RTqex2f	TTAGCCTCCCAAAGTGATGG	RT-PCR
SLC16A12RTqex4r	CGTTTGTGCGTAATCCTGAG	RT-PCR

Isolation and Cultivation of Vascular Smooth Muscle Cells

Vascular smooth muscle cells were isolated from radial arteries of patients undergoing coronary artery bypass grafting, as previously described.[28] Identification was made by immunofluorescent staining for smooth muscle α-actin (no. 1148818; Roche Diagnostics, Mannheim, Germany). Cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum and were used up to passage 10.

Statistical Analysis

DNA Analysis. For statistical calculation of odds ratio and significance, an open access Internet portal was used (http://faculty.vassar.edu/lowry/odds2x2.html; June 2009). Calculations were based on a $P = 0.05$ level of significance.

RNA Analysis. For the arithmetic mean of replicates (gDNA, $n = 4$; cDNA, $n = 6$), upper and lower confidence limits were determined using critical values of paired *t*-test distribution.

Mammalian Cell Culture Experiments. Averages and confidence intervals (paired *t*-test) were calculated and normalized on T-allele activity. Tissue culture experiments were performed twice using three technical replicas per measurement.

Results

Identification of Mutation in the 5′UTR of *SLC16A12* and Patient Characteristics

For mutation screening, we investigated genomic DNA of 350 patients with ARC and 134 patients with childhood cataract. DNA sequences from all exons and adjacent intron regions of *SLC16A12* were analyzed. A 79-year-old woman with ARC was found to carry a heterozygous sequence alteration in the 5′UTR (c.-17A>G; Fig. 1), but coding and approximately 50 nucleotides flanking intron sequences were unchanged. The sequence change was not found in any of the other 483 patients or in the 380 alleles of control subjects. This uniqueness makes the change unlikely to be a normal sequence variant; rather, it indicates pathogenicity. Hence, we will refer to it from now on as mutation. The patient was also heterozygous at an annotated SNP rs3740030 (c.-42T/G) that lies just upstream of the mutation. Cloning and sequencing of the patient's exon 3 showed that the mutation and the G-allele are on the same chromosome.

The patient was seen by an ophthalmologist for cataract. The lens of her left eye showed nuclear and subcapsular opacities plus pseudoexfoliation of the lens capsule that was removed when she was 75. At age 79, her right eye was also subjected to surgery. The right lens showed nuclear brunescent sclerosis with pulverulent anterior and posterior cortical opacities plus pseudoexfoliation of the lens capsule. No information about other family members was available.

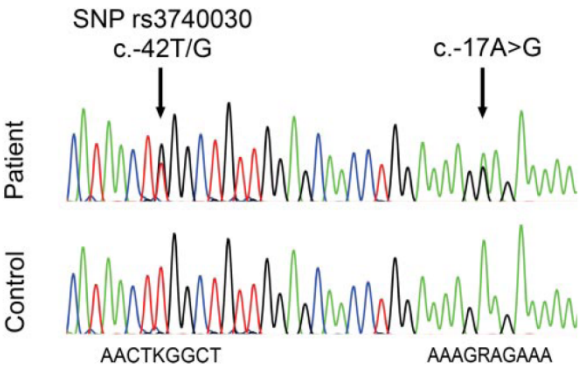

Figure 1. Electropherogram showing genomic DNA of 5′UTR section from control subject and ARC patient. Nucleotides of the immediate surroundings of SNP rs3740030 (c.-42) and the mutation (c.-17) are given. K = T and G; R = A and G.

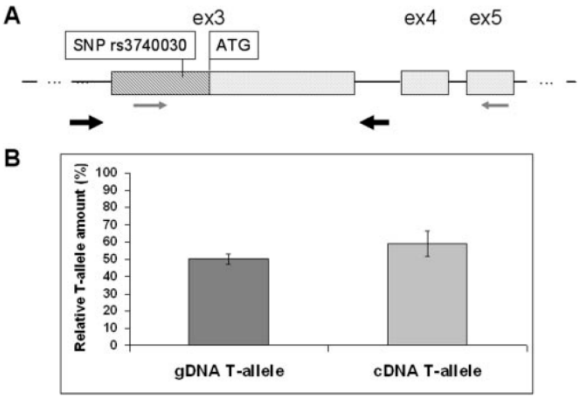

Figure 2. Quantitative sequencing analysis. (**A**) Drawing of exons (ex) 3, 4, and 5 of *SLC16A12*. Exon 3 contains the coding region beginning with ATG (*light gray box*) and 5′UTR (*hatched box*) containing SNP rs3740030. gDNA and transcript representing cDNA were obtained from vascular smooth muscle cells of a subject heterozygous for SNP rs3740030 (c.-42T/G). RT-PCR primers are indicated by *thin arrows* in exons 3 and exon 5. Locations of primers for genomic DNA amplification are indicated with *thick arrows* in intronic regions flanking exon 3. (**B**) Quantitative analysis expressed as relative amount of T-allele containing transcripts, given in percentages, for gDNA and mRNA (cDNA). Confidence intervals are shown for eight technical replicas.

Quantitative Assessment of *SLC16A12* Transcripts

Because the minor G-allele and the mutation reside within the 5′UTR of *SLC16A12*, it is likely that these changes affect translational rather than transcriptional efficiency. To assess the latter possibility, we determined the amount of *SLC16A12* mRNA by quantitative sequencing experiments, a method that requires RNA from heterozygous tissue. It is based on comparing allele frequencies of complementary DNA (cDNA) with those of gDNA from the same sample.[24] Given that no tissue was available from the cataract patient, we focused our attention on measuring SNP allele-specific transcripts in surrogate cells. Among available tissues known to express *SLC16A12*, vascular smooth muscle cells from heterozygous donors served for this purpose. RNA was subjected to one-step RT-PCR using primers in exons 3 and 5, to encompass the SNP (Fig. 2A). As expected, the contribution of the T-allele in genomic DNA was 50% (± 3%). No significant difference was measured for the T-allele in the cDNA (57% ± 9%; Fig. 2B), indicating that the two SNP alleles were present at comparable amounts. We concluded that regulatory mechanisms of translation rather than differential transcription were affected by the patient's sequence alterations.

Influence of *SLC16A12* 5′UTR Sequences at c.-42 and c.-17 on Translational Efficiency

To investigate whether the SNP and the patient's c.-17A>G mutation had effects on translation, constructs containing 5′UTR sequences in front of the luciferase gene were tested in HEK cells, by measuring luciferase reporter activity (Fig. 3). This assay is well suited to simulate translational activity

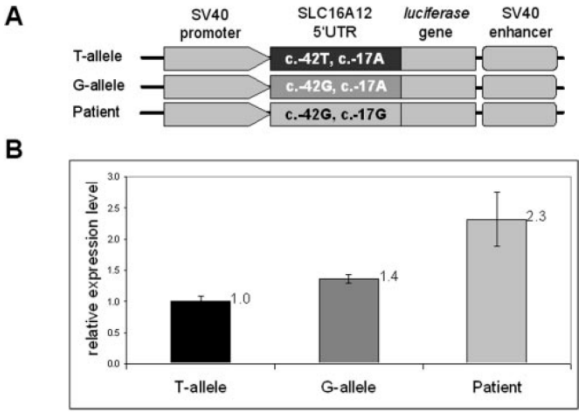

FIGURE 3. Luciferase reporter activity. (**A**) Schematic representation of cloned constructs. T-allele (c.-42T) and G-allele (c.-42G) contain 122 bp of *SLC16A12* 5′UTR. Patient construct contains c.-42G and c.-17G. (**B**) Relative luciferase activity. Values obtained for the T-allele were set to 1, to which other values were normalized. Displayed are the means (numerically) and confidence intervals (graphically) of three technical replicas. Results are from two independent experiments. The presence of the G-allele alone shows 1.4-fold and the presence of the G-allele in combination with the mutation (patient) shows 2.3-fold elevated luciferase expression levels.

changes because all constructs contain the same promoter, making it likely that transcriptional efficiency is the same among all constructs (Fig. 3A). Statistically significant differences in luciferase activity were observed among the three *SLC16A12* constructs (Fig. 3B). The construct that mimicked the patient's haplotype yielded the highest luciferase activity compared with both SNP alleles (2.315 ± 0.433-fold higher). Furthermore, the SNP alleles yielded significantly different luciferase activities such that the minor G-allele led to higher levels than did the more common T-allele (1.362 ± 0.07-fold higher). These results demonstrate a significant effect of 5′UTR sequences and suggest an enhanced potential of *SLC16A12* translation.

Effects of 5′UTR on RNA Folding by In Silico Analysis

One possible mechanism to achieve such regulation is through RNA folding. We simulated the influence of the 5′UTR sequence variants on secondary structures using the method Mfold,[27] which predicts a structure based on free energy minimization. RNA structures of the entire *SLC16A12* mRNA, (5′UTR, coding sequences, and 3′UTR) were predicted (Figs. 4A–C). Differences between the T- and G-alleles of SNP rs3740030 affect some branches near the centers of the molecules (Figs. 4A, 4B). The patient's mutation in combination with the SNP G-allele caused a strikingly different RNA structure (Fig. 4C). In addition, 5′UTR sequences alone folded in an allele-specific manner (Figs. 4D–F). Taken together, our experimental data, in combination with the bioinformatic predictions, indicate an effect on nucleotide-specific RNA folding properties.

SNP Allele-Specific Sequence Heterogeneity of 5′UTR *SLC16A12* Transcripts

The 5′UTR contains sequences from exons 1 and 2 and part of exon 3. To assess potential splicing effects on the generation of the 5′UTR, we initially applied the online tool ESEfinder 3.0. Splice factor SC35 was predicted to bind to the region immediately adjacent to the SNP, independently of the allele-specific nucleotide (Fig. 5). Interestingly, splice factor SRp40 was predicted to bind exclusively to G-allele transcripts and not to the T-allele transcripts. No binding sites for any of the ESEfinder 3.0–specific splice factors were found at the c.-17.

These observations supported the existence of alternative and allele-specific splicing of *SLC16A12* transcripts generating 5′UTR heterogeneity.

To test these predictions, we performed RT-PCR experiments with various primer combinations on cDNA from vascular smooth muscle cells of donors who were either homozy-

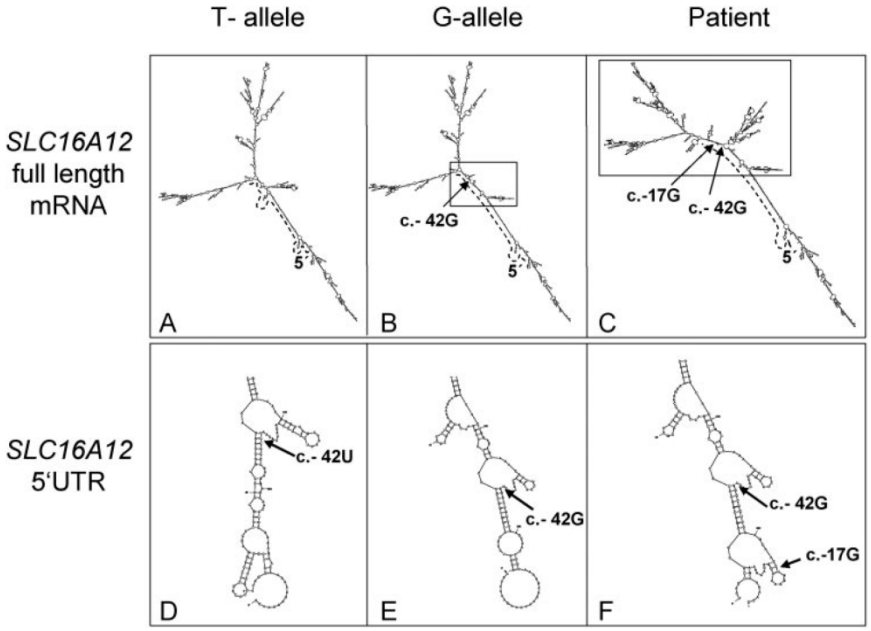

FIGURE 4. Predicted *SLC16A12* RNA foldings. T-allele and G-allele refer to the SNP rs3740030 (c.-42T/G). Patient has sequences c.-42G and c.-17G. (**A–C**) RNA foldings for the entire *SLC16A12* transcript. Differences in mRNA structure are highlighted (*box*). The 5′-end is marked (5′). *Dotted line*: 5′UTR. *Arrows*: positions c.-42G, c.-42U, and c.-17G. (**D–F**) Predicted RNA foldings for 5′UTR sequences only.

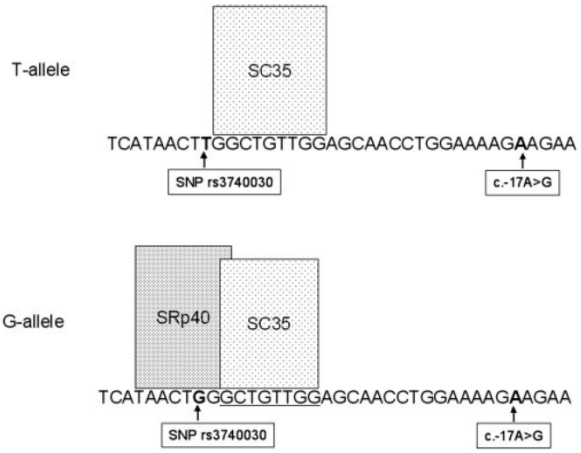

FIGURE 5. Prediction of splice factor binding site for *SLC16A12* 5′UTR. ESEfinder predictions were applied to cDNA from T- and G-alleles of SNP rs3740030. Factor SC35 binds independently of the allele sequence. In contrast, the binding of splice factor SRp40 is predicted only for the G-allele sequence. At the c.-17 position, no splice factor binding sites are predicted for G or A.

gous T/T or heterozygous G/T with respect to the SNP rs3740030 (Fig. 6). Primer combination III (exons 3–5) yielded amplicons of highly similar intensity (Fig. 6), confirming our quantitative transcript data that showed no allele-specific effect (Fig. 2). SNP nucleotide identity, T or G/T, was confirmed by sequencing. Amplification with primer combinations I yielded fragments displaying heterogeneities in size, sequence, and intensity. Surprisingly, in heterozygous samples, these were greatly reduced, and only two samples showed weakly staining single fragments (Fig. 6B). DNA sequence analysis of sample 5 (primer combination II) revealed the expected presence of exons 2, 3, and 4 but unexpectedly only the T nucleotide of the SNP. Sample 8 (primer combination I) contained a short fragment intron 1 sequence in addition to exons 1, 2, and 3. At the SNP position, only the G-nucleotide was found. The weak staining of this fragment suggested reduced levels of this transcript, which may be the reason for chance amplification. Sequence heterogeneity in the homozygous samples resulted from some fragments containing exons 1, 2, and 3, some had an additional short segment of intron 1, and yet others lacked exon 2. Primer combination II amplified fragments containing sequences from exons 2, 3, and 4 (DNA sequence data not shown). These striking differences between the homozygous and heterozygous samples within the 5′UTR region indicated the possibility of an allele-specific splicing mechanism that led to severe reductions in transcripts containing exons 1 and 2 in heterozygous tissues.

Affect of Allele Specificity on the Risk for Cataract

Assessment of the genotype for SNP rs3740030 in 484 cataract patients (350 with ARC, 134 with childhood cataract), most of whom were of Caucasian background, and in 190 ethnically matched control subjects confirmed an overrepresentation of the T-allele (Table 2), as had been reported for Caucasian populations of European descent (HapMap: T-allele frequency of 91.6%). Furthermore, we noticed overrepresentation of the G-allele in patient groups compared with control subjects (juvenile cataract, 6.72%; ARC, 8.71%, control subjects, 4.21%) that was statistically significant only for the ARC patient group ($P = 0.00601768$; $\chi^2 = 7.545$). To assess the extent to which the G-allele could increase the risk for cataract, odds ratios were calculated. Persons with either the G-allele only or the G/T or G/G genotype are at an approximately 2.2-fold increased risk for ARC (Table 3). Conversely, the T-allele or the T/T genotype confers protection against ARC. Given that the control population originated from the general Swiss population, we cannot exclude the possibility that an individual of this control population may develop ARC. In such case, we would predict that the odds ratio shifts even further toward a risk association with the G-allele. Odds ratios for juvenile cataract were statistically not significant (Table 3).

DISCUSSION

The complex disease ARC is influenced by a multitude of environmental and genetic factors. Knowledge of the identity of the genetic contributions is limited and focuses on the identification of rare variants or the association of sequence variants in affected families and patient populations[11–14,16,18] and on whole genome scans.[29] Previously, we showed that a mutation in the monocarboxylate transporter *SLC16A12* causes juvenile cataract, likely through a disturbance of solute homeostasis.[6] Given that maintenance of homeostasis within the aging lens is likely to be essential,[8] this transporter is a prime candidate for ARC. The data presented here support this hypothesis. First, a point mutation and an SNP within the 5′UTR was found in a 79-year-old woman with ARC. Second, in vitro and ex vivo experiments provide evidence that these sequences affect expression of *SLC16A12* by modulating translational efficiency. Third, the SNP within the 5′UTR is likely associated with ARC within a Swiss population.

The c.-17A>G alteration in the patient with ARC was unique among our patient cohort ($n = 484$) and was also absent in our control population ($n = 180$), suggesting that it is a mutation rather than an SNP. Supportively, no sequence variation at position c.-17 of *SLC16A12* has been reported in

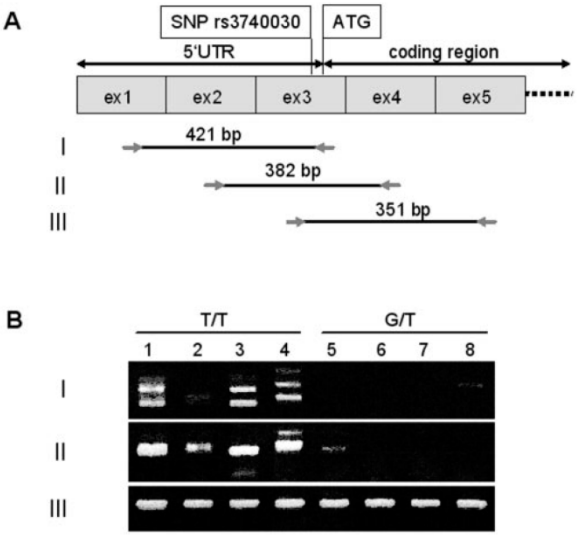

FIGURE 6. *SLC16A12* transcript analysis by RT-PCR. (**A**) Schematic representation of exons 1 through 5 (size not to scale). Position of SNP rs3740030 and the ATG initiation codon as well as the 5′UTR and coding regions are indicated above the exons. Primer combinations (I, II, III) and their respective sizes (bp) are shown. (**B**) Electrophoretically separated RT-PCR products from homozygous T/T (1–4) and heterozygous G/T (5–8) donors of vascular smooth muscle cells.

TABLE 2. Genotype and Allele Frequencies of SNP rs3740030

	Genotypes				Alleles				
	TT	TG	GG	Total	T	G	Total	T (%)	G (%)
Control	175	14	1	190	364	16	380	95.79	4.21
Juvenile	116	18	0	134	250	18	268	93.28	6.72
Age-related	291	57	2	350	639	61	700	91.29	8.71

Number of subjects identified with respective genotypes and alleles. Allele frequencies are given in percentages.

available DNA sequence databases to date. The second alteration found *in cis* in the age-related patient is the only annotated exonic SNP rs3740030 (c.-42T/G). Its allele frequency is population dependent. In contrast to a Caucasian population of European descent in which the minor allele is relatively seldom (8.4%), its frequency is higher in a Han Chinese population (27.4%) (HapMap, January 2010). It would be intriguing to investigate whether this is related to the incidence rate of cataract in the Chinese population.[30] Based on our data in a Swiss population, the minor G-allele was more frequent in ARC patients than in the control group. This increased frequency translates into an approximately 2.2-fold increased risk and renders SNP rs3740030 a potential modifier in the development of ARC. Numerous examples from genome-wide association studies demonstrate a population-specific effect. ARC is no exception. For example, variants of GST have been investigated in various ethnically different cataract patient populations with varying results. Although no association was found in an Italian population,[17] the opposite was true for an Estonian[11] and a Turkish[10] population. The SNP in *SLC16A12* studied here may underlie similar conditions because the frequency in the Swiss control population was lower than that reported by HapMap. In addition, in preliminary studies, we were unable to verify the association in a British control population.

Based on the location of the opacity within the lens, one distinguishes between nuclear (the most frequent form), cortical, posterior subcapsular, and mixed-type cataract.[31] Some risk factors have been reported as associated specifically with one of the subtypes of ARC,[10,11,13] whereas association of the N-acetyltransferase 2 isoform "slow" acetylator[12] and of kinesin light-chain *KCL1* sequence variants[32] have not been. We did not differentiate between subtypes; therefore, the association of SNP rs3740030 in *SLC16A12* could account for all ARC in our study, but we cannot exclude that the increased risk would also be subtype specific.

The underlying molecular basis for ARC can be explained by altered expression levels of *SLC16A12*. Luciferase activity increases from the common T-allele to the minor G-allele to the patient haplotype (G-allele and mutation), thus offering an explanation for predisposition to and manifestation of the disease. Most likely the pathogenic mechanism acts posttranscriptionally because equal amounts of the SNP allele-specific *SLC16A12* transcripts were found and 5′UTR regions are known to harbor control elements that regulate translation. This regulation may occur by way of RNA folding, upstream open-reading frames (uORFs), differential splicing, or specific sequence elements. We found no evidence for the latter. In contrast, the RNA-folding models demonstrated visible allele-specific differences, with a most drastic change for the patient allele. These structural changes correlate with changes in luciferase reporter activity. In all likelihood, no single mechanism is responsible for the observed effects. In addition, or complementary, to RNA folding, uORFs may be involved. Alterations in uORFs of other genes (*TPO*, *HTR3A*, *BRAC1*, *TGFß3*) were reported to cause various diseases by enhancing gene expression through elevation of the level of translation.[22,33] The *SLC16A12* 5′UTR contains four uORFs. One of them, which is not in frame with *SLC16A12*, harbors both sequence alterations discussed here. Given that this uORF expands into the coding frame, it may have an inhibitory effect on guiding ribosomes beyond the initiation of the ATG codon.[34–43] The additional feature that the SNP G-allele changes the coding frame of that uORF slightly (p.L21W) could also influence translational efficiency.

Differential splicing of the 5′UTR may also explain the allele-specific effects. Examples have been reported for the TFII-I transcription factor[44] and for the estrogen receptor β gene.[45] For the SNP rs3740030, the action of splice factor SRp40, which is predicted to bind exclusively to transcripts generated from the minor G-allele, may cause the observed allele-specific 5′UTR sequence heterogeneity. Furthermore, the skipping of exon 2 causes shortening of the 5′UTR and elimination of another uORF, thereby increasing the likelihood of influencing the translational regulation. Taken together, at least three different mechanisms may apply, singly or in concert, to explain the altered translational efficiency associated with the SNP and the mutation within the 5′UTR.

Knowledge of the exact function of the monocarboxylate transporter SLC16A12 requires identification of its substrates, which is currently under investigation. Nevertheless, the *SLC16A12* nonsense mutation leads to juvenile cataract[6] and nucleotide changes in the 5′UTR associate with ARC, demonstrating the importance of this transporter. The adult lens

TABLE 3. Association of SNP rs3740030 with Cataract

	Allele					Genotype				
	Odds Ratio	CI		*P*	χ^2	Odds Ratio	CI		*P*	χ^2
Juvenile	1.638	0.8196	3.2736	0.158341	1.99	1.8103	0.8774	3.7352	0.104862	2.63
Age-related	2.1718	1.2339	3.8225	0.006001	7.55	2.3654	1.3021	4.297	0.003836	8.36

Odds ratio, confidence interval (CI), *P*, and χ^2 are shown. Statistical significance was reached if odds ratio and CI were >1, $P < 0.05$, and $\chi^2 > 3.84$. Analyzed were 190 control subjects, 134 patients with juvenile cataract, and 350 patients with age-related cataract.

structure demands a well-functioning transport system for metabolites to supply nutrients and to remove waste products,[8] and the solute carrier SLC16A12 may have an important function in that activity. Complete removal of the transporter leads to congenital cataract and renal glucosuria.[6] In contrast, alterations in the amount of the transporter, as described here, may lead to imbalanced solute homeostasis, which over time leads to ARC. Given *SLC16A12* expression in tissues other than the lens, it seems feasible that other organs could experience such consequences. Renal glucosuria is a prime candidate of such a condition but other, as yet not described, physiological conditions may become manifest.

We have provided initial examples that *SLC16A12* 5′UTR alterations are involved in the manifestation of and the predisposition to ARC by modifying translational efficiency. Physiological consequences are likely to challenge tissue homeostasis and to transform a healthy lens into a diseased lens.

Acknowledgments

The authors thank Helen Greutert for help with cell cultures, Walther Hänseler for help with cloning, and Gabor Matyas for sharing his knowledge of DNA sequencing and for maintaining the sequencing facility.

References

1. Brian G, Taylor H. Cataract blindness—challenges for the 21st century. *Bull World Health Organ.* 2001;79:249–256.
2. Congdon NG, Friedman DS, Lietman T. Important causes of visual impairment in the world today. *JAMA.* 2003;290:2057–2060.
3. Shiels A, Hejtmancik JF. Genetic origins of cataract. *Arch Ophthalmol.* 2007;125:165–173.
4. Francis PJ, Berry V, Bhattacharya SS, Moore AT. The genetics of childhood cataract. *J Med Genet.* 2000;37:481–488.
5. Shiels A, Bennett TM, Knopf HL, et al. CHMP4B, a novel gene for autosomal dominant cataracts linked to chromosome 20q. *Am J Hum Genet.* 2007;81:596–606.
6. Kloeckener-Gruissem B, Vandekerckhove K, Nurnberg G, et al. Mutation of solute carrier SLC16A12 associates with a syndrome combining juvenile cataract with microcornea and renal glucosuria. *Am J Hum Genet.* 2008;82:772–779.
7. Jamieson RV, Farrar N, Stewart K, et al. Characterization of a familial t(16;22) balanced translocation associated with congenital cataract leads to identification of a novel gene, TMEM114, expressed in the lens and disrupted by the translocation. *Hum Mutat.* 2007;28:968–977.
8. Truscott RJ. Age-related nuclear cataract: a lens transport problem. *Ophthalmic Res.* 2000;32:185–194.
9. Hodge WG, Whitcher JP, Satariano W. Risk factors for age-related cataracts. *Epidemiol Rev.* 1995;17:336–346.
10. Guven M, Unal M, Sarici A, Ozaydin A, Batar B, Devranoglu K. Glutathione-S-transferase M1 and T1 genetic polymorphisms and the risk of cataract development: a study in the Turkish population. *Curr Eye Res.* 2007;32:447–454.
11. Juronen E, Tasa G, Veromann S, et al. Polymorphic glutathione S-transferases as genetic risk factors for senile cortical cataract in Estonians. *Invest Ophthalmol Vis Sci.* 2000;41:2262–2267.
12. Tamer L, Yilmaz A, Yildirim H, et al. N-acetyltransferase 2 phenotype may be associated with susceptibility to age-related cataract. *Curr Eye Res.* 2005;30:835–839.
13. Shiels A, Bennett TM, Knopf HL, et al. The EPHA2 gene is associated with cataracts linked to chromosome 1p. *Mol Vis.* 2008;14:2042–2055.
14. Bhagyalaxmi SG, Srinivas P, Barton KA, et al. A novel mutation (F71L) in αA-crystallin with defective chaperone-like function associated with age-related cataract. *Biochim Biophys Acta.* 2009;1792:974–981.
15. Maraini G, Hejtmancik JF, Shiels A, et al. Galactokinase gene mutations and age-related cataract: lack of association in an Italian population. *Mol Vis.* 2003;9:397–400.
16. Okano Y, Asada M, Fujimoto A, et al. A genetic factor for age-related cataract: identification and characterization of a novel galactokinase variant, 'Osaka,' in Asians. *Am J Hum Genet.* 2001;68:1036–1042.
17. Alberti G, Oguni M, Podgor M, et al. Glutathione S-transferase M1 genotype and age-related cataracts: lack of association in an Italian population. *Invest Ophthalmol Vis Sci.* 1996;37:1167–1173.
18. Shi Y, Shi X, Jin Y, et al. Mutation screening of HSF4 in 150 age-related cataract patients. *Mol Vis.* 2008;14:1850–1855.
19. Hansen L, Mikkelsen A, Nurnberg P, et al. Comprehensive mutational screening in a cohort of Danish families with hereditary congenital cataract. *Invest Ophthalmol Vis Sci.* 2009;50:3291–3303.
20. Reynolds PR. In sickness and in health: the importance of translational regulation. *Arch Dis Child.* 2002;86:322–324.
21. Scheper GC, van der Knaap MS, Proud CG. Translation matters: protein synthesis defects in inherited disease. *Nat Rev Genet.* 2007;8:711–723.
22. Chatterjee S, Pal JK. Role of 5′- and 3′-untranslated regions of mRNAs in human diseases. *Biol Cell.* 2009;101:251–262.
23. Cazzola M, Skoda RC. Translational pathophysiology: a novel molecular mechanism of human disease. *Blood.* 2000;95:3280–3288.
24. Magyar I, Colman D, Arnold E, et al. Quantitative sequence analysis of FBN1 premature termination codons provides evidence for incomplete NMD in leukocytes. *Hum Mutat.* 2009;30:1355–1364.
25. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: a Web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* 2003;31:3568–3571.
26. Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet.* 2006;15:2490–2508.
27. Zuker M. Mfold Web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003;31:3406–3415.
28. Weiss S, Frischknecht K, Greutert H, et al. Different migration of vascular smooth muscle cells from human coronary artery bypass vessels: role of Rho/ROCK pathway. *J Vasc Res.* 2007;44:149–156.
29. Iyengar SK, Klein BE, Klein R, et al. Identification of a major locus for age-related cortical cataract on chromosome 6p12–q12 in the Beaver Dam Eye Study. *Proc Natl Acad Sci U S A.* 2004;101:14485–14490.
30. Hsu WM, Cheng CY, Liu JH, Tsai SY, Chou P. Prevalence and causes of visual impairment in an elderly Chinese population in Taiwan: the Shihpai Eye Study. *Ophthalmology.* 2004;111:62–69.
31. Klein BE, Klein R, Linton KL. Prevalence of age-related lens opacities in a population: the Beaver Dam Eye Study. *Ophthalmology.* 1992;99:546–552.
32. Andersson ME, Zetterberg M, Tasa G, et al. Variability in the kinesin light chain 1 gene may influence risk of age-related cataract. *Mol Vis.* 2007;13:993–996.
33. Hughes TA. Regulation of gene expression by alternative untranslated regions. *Trends Genet.* 2006;22:119–122.
34. Alderete JP, Jarrahian S, Geballe AP. Translational effects of mutations and polymorphisms in a repressive upstream open reading frame of the human cytomegalovirus UL4 gene. *J Virol.* 1999;73:8330–8337.
35. Beffagna G, Occhi G, Nava A, et al. Regulatory mutations in transforming growth factor-β3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. *Cardiovasc Res.* 2005;65:366–373.
36. Degnin CR, Schleiss MR, Cao J, Geballe AP. Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. *J Virol.* 1993;67:5514–5521.
37. Hill JR, Morris DR. Cell-specific translational regulation of S-adenosylmethionine decarboxylase mRNA: dependence on translation and coding capacity of the cis-acting upstream open reading frame. *J Biol Chem.* 1993;268:726–731.
38. Mize GJ, Ruan H, Low JJ, Morris DR. The inhibitory upstream open reading frame from mammalian S-adenosylmethionine decarboxylase mRNA has a strict sequence specificity in critical positions. *J Biol Chem.* 1998;273:32500–32505.

39. Niesler B, Flohr T, Nothen MM, et al. Association between the 5′ UTR variant C178T of the serotonin receptor gene HTR3A and bipolar affective disorder. *Pharmacogenetics.* 2001;11: 471–475.
40. Parola AL, Kobilka BK. The peptide product of a 5′ leader cistron in the β2 adrenergic receptor mRNA inhibits receptor synthesis. *J Biol Chem.* 1994;269:4497–4505.
41. Wang Z, Fang P, Sachs MS. The evolutionarily conserved eukaryotic arginine attenuator peptide regulates the movement of ribosomes that have translated it. *Mol Cell Biol.* 1998;18:7528–7536.
42. Werner M, Feller A, Messenguy F, Pierard A. The leader peptide of yeast gene CPA1 is essential for the translational repression of its expression. *Cell.* 1987;49:805–813.
43. Morris DR, Geballe AP. Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol.* 2000;20:8635–8642.
44. Makeyev AV, Bayarsaihan D. Alternative splicing and promoter use in TFII-I genes. *Gene.* 2009;433:16–25.
45. Smith L. Post-transcriptional regulation of gene expression by alternative 5′-untranslated regions in carcinogenesis. *Biochem Soc Trans.* 2008;36:708–711.

Contributions of authors to the manuscript "Alterations of the 5'Untranslated Leader Region of *SLC16A12* Lead to Age-related Cataract"

Jurian Zuercher	Planning and performing experiments, preparation of clones, primer design, interpretation of data, writing of the manuscript
John Neidhardt	Preparation of RNA, RT-PCR, editing of the manuscript
István Magyar	Quantitative transcript analysis, editing of the manuscript
Stephan Labs	Providing of technical help, editing of the manuscript
Anthony T. Moore	Conceptual input, sending patient`s material, editing of the manuscript
Felix C. Tanner	Providing of cells for cell culture experiments, editing of the manuscript
Naushin Waseem	Conceptual input, sending patient`s material, editing of the manuscript
Daniel F. Schorderet	Sending patient`s material, editing of the manuscript
Francis L. Munier	Sending patient`s material, editing of the manuscript
Shomi Bhattacharya	Conceptual input, sending patient`s material, editing of the manuscript
Wolfgang Berger	Conceptual input, editing of the manuscript
Barbara Kloeckener-Gruissem	Conceptual planning, design and supervision of the study, interpretation of data, writing and editing of the manuscript

Appendix 3

Primer sequences, combinations, and conditions used in this work

Oligo Name	Sequence (5`-3`)	Oligo Length (bp)	Tm (ºC)	Additional Information		Product Length (bp)
				Q-Solution	Number of Cycles	
FBN1_5UTR_F	CTC GGG GAT TTG TCT CTG TGT	21	58	+	40	417
FBN1_N210_R	TCC AGG GCA ACA GTA AGC ATT A	22				
FBN1_N87_F	GGA GGC TGG GAA CGT GAA G	19	64	+	35	257
FBN1_N323_R	ATC TGG AGC CAC AGG AAG GAG	21				
FBN1_N87_F	GGA GGC TGG GAA CGT GAA G	19	58	+	35	398
FBN1_N584_R	GCC ACA CAC CTT CCT CCA TTG	21				
FBN1_N361_F	CCT GTG GGG ATG GAT TTT GTT C	22	64	-	40	223
FBN1_N584_R	GCC ACA CAC CTT CCT CCA TTG	21				
FBN1_N418_F	ACA TAG GGA CTC ACT GTG GAC AAC	24	58	+	35	488
FBN1_N906_R	ACC CCC TTC ACA GAT TCC AG	20				
FBN1_N655_F	TGG ACC CCA GTG TGA AAG AGA TT	23	58	-	35	237
FBN1_N888_R	GCC TGG CAT TCA TCC ACA TC	20				
FBN1_N813_F	AAA TGC CCT GCT GGA CAC AA	20	64	-	35	334
FBN1_N1147_R	CGG TTG CTC TGA TGG GAC AC	20				
FBN1_N1060_F	CCA AAA TGC AGT GCT GCT GTG	21	62	+	40	269
FBN1_N1329_R	CCT TGG TGG CTC CCG AGA T	19				
FBN1_N1096_F	GGT CTC CAG GGG TCA CTG TC	20	62	-	40	434
FBN1_N1530_R	CGA ACC CTG GTT GTT AAT ACA CTC A	25				
FBN1_N1238_F	CCC TGT TCC TCC TGG CTT T	19	62	-	40	408
FBN1_N1646_R	GTG TTG ATG CAG CGT CCA TTA T	22				
FBN1_N1447_F	ACC TCC GTG GGG AGT GTA TT	20	62	-	35	312
FBN1_N1759_R	ACA TTC CAT TAA GGC ACA TGT TC	23				
FBN1_N1321_F	CAC CAA GGG TGC TGC CAG TA	20	66	-	35	231
FBN1_N1552_R	CAG CTC GGC ACT GAC AGG TGT A	22				
FBN1_N1321_F	CAC CAA GGG TGC TGC CAG TA	20	58	+	35	438
FBN1_N1759_R	ACA TTC CAT TAA GGC ACA TGT TC	23				
FBN1_N1386_F	GGA CGC TGC ATT CCA ACT C	19	58	+	35	280
FBN1_N1666_R	CGC AAT GAA AAC TGC CAT CT	20				
FBN1_N1447_F	ACC TCC GTG GGG AGT GTA TT	20	62	-	35	312
FBN1_N1759_R	ACA TTC CAT TAA GGC ACA TGT TC	23				
FBN1_N1579_F	AAT GCC GAG ACA TTG ATG AGTG	22	62	-	35	345

FBN1_N1861_R	CAG GGA AGC ATT CAC ATC TGT AG	23				
FBN1_N1838_F	CAT GCG GAG CAC ATG CTA T	19	58	-	35	412
FBN1_N2379_R	GGG GCA GGT ACA GAC AAA ACT TC	23				
FBN1_N1945_F	GCT GGC ATC AGA TGG ACG TTA T	22	62	+	40	250
FBN1_4_R	GGC GCA ACA GCA TTC AGA TT	20				
FBN1_N2157_F	GCA GGC AGT GAT ATA AAT GAA TGT G	25	58	+	40	319
FBN1_N2456_R	CTG GGC TGT TCT TGC AGA CTC	21				
FBN1_N2430_F	TGC GAA TCA AGT CCT TGC ATT A	22	58	+	40	405
FBN1_N2835_R	GGC ATC CAA AGT CAT TCC ACT G	22				
FBN1_N2678_F	TCC CAT ATG TGG TAA AGG GTA CTC A	25	58	-	35	477
FBN1_N3155_R	TTA AAG CTG CCA ATG GTG TTT CT	23				
FBN1_N2808_F	GTG TCC CAG TGG AAT GAC TTT G	22	58	+	40	349
FBN1_N3156_R	CTT AAA GCT GCC AAT GGT GTT TC	23				
FBN1_N2977_F	GCG AGG AGT GTC CCA TGA	18	58	+	40	270
FBN1_N3247_R	TGC CAC AGA GGT CAG GAG ATA	21				
FBN1_N2977_F	GCG AGG AGT GTC CCA TGA G	19	66	+	40	247
FBN1_N3224_R	CGG CAT TCG TCA ATG TCT GT	20				
FBN1_N3192_F	GAA AGG AAC TGC ACA GAC ATT GA	23	58	+	35	230
FBN1_N3422_R	GGG CAT TCA CAG CGG TAA C	19				
FBN1_N3451_F	CCG CGT GTA TCG ACA TCA A	19	58	-	40	228
FBN1_N3679_R	ATC CCG GCT GAC AGC TAC AT	20				
FBN1_N3451_F	CCG CGT GTA TCG ACA TCA A	19	58	-	40	653
FBN1_N4080_R	GGA ACA TTC GTC CAG ATC AGT G	22				
FBN1_N3600_F	TGC AGC ATA ATG AAT GGT GGT TGT	24	58	+	40	403
FBN1_N4003_R	TGC CAC AGT TGT GTG CTC CA	20				
FBN1_N4111_F	CCC ATA TGT GCA GCC AGC AT	20	66	-	35	252
FBN1_N4343_R	TGT TCG GAA GGG AGC ACT CA	20				
FBN1_N3870_F	TGC CTA AGT GGG ACC TGT GA	20	62	-	35	269
FBN1_N4080_R	GGA ACA TTC GTC CAG ATC AGT G	22				
FBN1_N3870_F	TGC CTA AGT GGG ACC TGT GA	20	58	+	40	493
FBN1_N4343_R	TGT TCG GAA GGG AGC ACT CA	20				
FBN1_N3892_F	CAA CTG TGG CAA ACA TGC TGT A	22	58	-	35	237
FBN1_N4229_R	TCA GAG CAC TCA TCA AGG TCT GT	23				
FBN1_N3668_F	TCA GCC GGG ATT TGC ACT A	19	62	+	40	335
FBN1_N4003_R	TGC CAC AGT TGT GTG CTC CA	20				
FBN1_N3688_F	TCA GCC GGG ATT TGC ACT A	19	58	-	35	215
FBN1_N3903_R	TGA GCC TTT CGT GTT TTC ACA G	22				
FBN1_N3806_F	CAT GGC ATC TGA AGA CAT GAA GA	23	58	-	35	398

FBN1_N4204_R	AAG TGA AGC CAT CAC CTG TGT ATC	24				
FBN1_N4128_F	CAT GCA GAC TGC AAG AAT ACC A	22	62	-	35	234
FBN1_N4362_R	GTT CGG AAG GGA GCA CTC A	19				
FBN1_N4259_F	CCT CAA TGC ACC AGG AGG ATA C	22	58	+	40	445
FBN1_N4704_R	TTT ACC CAG AGA ACA GCA GCA G	22				
FBN1_N4318_F	ACG GGA AAG CCT GTG AAG ATA	21	58	+	40	230
FBN1_N4548_R	ATC AGG TGG GCA GTC ACA GA	20				
FBN1_N4533_F	GAC TGC CCA CCT GAT TTT GAA	21	58	-	35	246
FBN1_N4779_R	TTC CCC TCC AGG ACA AAG AAT	21				
FBN1_N4682_F	CTG CTG CTG TTC TCT GGG TAA A	22	58	-	40	305
FBN1_N4987_R	ATG TCC CTG GAC CAC AGA TTC	21				
FBN1_N5045_F	TGG GGG AAA TAA TTG CAT GG	20	58	-	35	273
FBN1_N5318_R	ATC TCC CGG CAC TCA TCA ATA	21				
FBN1_N5186_F	GAA CAA GCC CTG TGA ACA GTG TC	23	58	-	35	306
FBN1_N5492_R	TAG CTG CCT GCA GTG TTG ATG	21				
FBN1_N5256_F	CCA GGC TTT GTC ATC GAC AT	20	64	-	35	289
FBN1_N5545_R	CAT TGC ACT GTC CTG TGG AG	20				
FBN1_N5461_F	ACG CCG AAT GCA TCA ACA CT	20	62	-	40	360
FBN1_N5820_R	GAT TCC CAT TTC CAC TTG CAC AT	23				
FBN1_N5620_F	GCC TTT GCC ACA CTG GTT TT	20	66	-	35	255
FBN1_N5875_R	TGC ACT GGC ACT GGA AAG AC	20				
FBN1_N5655_F	CAA ACC ATG TGC TTG GAC ATA A	22	62	-	35	302
FBN1_N5957_R	GGT GCA CAT TTT CTG GGT TCT A	22				
FBN1_N5817_F	AAA TGG GAA TCT TTG CAG AAA TG	23	58	-	40	257
FBN1_N6067_R	TTT CTG GCT CTT CGA CAC ACT C	22				
FBN1_N5970_F	AAC TTG GAT GGG TCC TAC AGA TG	23	58	+	40	279
FBN1_N6249_R	CAT TCC TGC TTG GAG TGA TTT C	22				
FBN1_N6212_F	ATC ACC CAA ATC CAG AAA TCA C	22	58	+	40	310
FBN1_N6522_R	ATT GCC AAC AGA ACA TTC ATC AG	23				
FBN1_N6213_F	TCA CCC AAA TCC AGA AAT CAC TC	23	58	-	40	366
FBN1_N6579_R	CTC GCA GGT GCA TTC AAA AC	20				
FBN1_N6515_F	TGG CAA TCC TTG TGG AAA TG	20	62	+	40	295
FBN1_N6810_R	GCC AAT GAG GTT CTT GCA TTC	21				
FBN1_N6666_F	AAG ACC GTA GGA TGT GCA AAG AT	23	58	+	40	355
FBN1_N7073_R	ACG GGG TTC CTG TTG CTG	18				
FBN1_N6841_F	AGC GGA GAC CTG ATG GAG AA	20	62	+	40	404
FBN1_N7245_R	ATT TCG GCA AAC ATC GTG AAT AA	23				
FBN1_N6911_F	TGG GCG CTG CCT CAA C	16	64	+	40	447

FBN1_N7358_R	TTG GGA GCC TGG TTG CAC T	19				
FBN1_N6974_F	CCC CAA CCA GGA CGA GTG	18	58	+	40	269
FBN1_N7243_R	TTC GGC AAA CAT CGT GAA TAA C	22				
FBN1_N7141_F	AGG GGA CTG TGG CTT TCA AGA	21	62	+	40	381
FBN1_N7522_R	TGA AGC CGC CAA TGG TGT T	19				
FBN1_N7408_F	GCC CGA AAG GCT ACA TTC TG	20	62	-	40	321
FBN1_N7730_R	AGC GGT GGT TAC CCT CAC A	19				
FBN1_N7077_F	AAA TCG GAA TGC TGC TGT GA	20	58	+	40	272
FBN1_N7349_R	TGG TTG CAC TCG TTC AGA TCT AC	23				
FBN1_N7681_F	CCG GCT CCA GCT GTG AAG	18	62	-	40	458
FBN1_N8139_R	CTC TGG GGA GAG TGA ATT GTC AT	23				
FBN1_N7732_F	AGC ATG GCT GCC AGA ACA	18	58	+	35	342
FBN1_N8075_R	TGC CCA TTC CAG AAA CAC AG	20				
FBN1_N8201_F	CGA AAC TGA TGC CTC CAA TAT C	22	58	+	40	435
FBN1_N8636_R	TTT GGT CTC TGG ATG GTG AAT TA	23				
SM_FBN1_F	CTG GGG ACT TTG AAT GCA AGT G	22	58	+	40	286
SM_FBN1_R	CCA GGG TTG CAG GCA CAC	18				
FBN1_3UTR_F	CAG GCT TCT CAT GTG TGT AGC TAA	24	58	-	35	313
FBN1_3UTR_R	ATT GGA ACT GAA TAC TTG TAA TTT GGT	27				
FBN1_SB_Ex52_F	CTG CGA GTG TCC CTT TGG TTA T	22	58	-	35	361
FBN1_SB_Ex55_R	GCC AAT GAG GTT CTT GCA TTC	21				
FBN1_Ex15_F	GCG TTG CGT CAA CAC TGA T	19	62	+	35	267
FBN1_Ex17_R	GGC CCA CTG CTG CAG AG	17				
FBN1_Ex52_F	CGC TGC GAG TGT CCC TTT	18	58	+	40	255
FBN1_Ex54_R	CAC GGG ACA TTT GCA TTC A	19				
FBN1_N4259_F	CCT CAA TGC ACC AGG AGG ATA C	22	58	+	40	289
FBN1_N4548_R	ATC AGG TGG GCA GTC ACA GA	20				
FBN1_N4976_F	GTC CAG GGA CAT GTT ACA ACA	21	58	+	40	264
FBN1_N5240_R	GGC CTT TGA CTT CCA CAG A	19				
FBN1_N1487_F	AAA ACC CCT GTG CTG GTG	18	58	+	40	307
FBN1_N1794_R	CAG CTG GAA TCC AGG TTT G	19				
FBN1_Ex63_F	TGG CTG TCC ACC TGG TTA CT	20	58	+	40	423
FBN1_Ex65_R	CTG ATC CCT TCC TTT TGG TTG A	22				
FBN1_N3892_F	CAA CTG TGG CAA ACA TGC TGT A	22	58	+	40	397
FBN1_N4390_R	GTT GTG GCA AGT TCC AAA GAC A	22				

Appendix 4

Curriculum Vitae

PERSONAL DATA

Name: István Magyar
Date of birth: October 12, 1979
Place of birth: Szentes (Hungary)
Nationality: Hungarian

EDUCATION

01/06 – present **University of Zurich, Switzerland - Promotion**
Institute of Medical Molecular Genetics
Title of Dissertation: Molecular Basis of Marfan Syndrome: *In vitro* and *in silico* Analyses of Exonic and Intronic Sequence Variants in the *FBN1* Gene

09/00 – 05/05 **Szent István University, Hungary – MSc Diploma**
Specialization in Molecular Biology and Biotechnology
Title of Diploma Thesis: Identification of the *glmS* Gene in *Azospirillum* Species

09/94 – 06/98 **Gregus Mátè Secondary School, Hungary - High School Diploma**

PAPERS

Magyar I, Arnold E, Colman D, Meienberg J, Baumgartner D, Bottani A, Fokstuen S, Addor M-C, Carrel T, Steinmann B, Berger W, Mátyás G. (2011) *FBN1* pre-mRNA Splicing Alterations: Comparative *in vitro* Transcript Analyses and *in silico* Prediction for the Effect of *FBN1* Exonic and Intronic Mutations (Manuscript in preparation).

Zürcher J, Neidhardt J, **Magyar I**, Labs S, Moore AT, Tanner FC, Waseem N, Schorderet DF, Munier FL, Bhattacharya S, Berger W, Kloeckener-Gruissem B. (2010) Alterations of the 5'Untranslated Leader Region of *SLC16A12* Lead to Age-related Cataract. Invest Ophthalmol Vis Sci 51(7):3354-3361.

Magyar I, Colman D, Arnold E, Baumgartner D, Bottani A, Fokstuen S, Addor M-C, Berger W, Carrel T, Steinmann B, Mátyás G. (2009) Quantitative Sequence Analysis of *FBN1* Premature Termination Stop Codons Provide Evidence for Incomplete NMD in Leukocytes. Hum Mutat 30(9):1355-1364.

Mátyás G, Alonso S, Patrignani A, Marti M, Arnold E, **Magyar I**, Henggeler C, Carrel T, Steinmann B, Berger W. (2007) Large Genomic Fibrillin-1 (*FBN1*) Gene Deletions Provide Evidence For True Haploinsufficiency in Marfan Syndrome. Hum Genet 122:23-32.

Banin E, Mizrahi-Meissonnier L, Neis R, Silverstein S, **Magyar I**, Abeliovich D, Roepman R, Berger W, Rosenberg T, Sharon D. (2007) A Non-Ancestral *RPGR* Missense Mutation in Families With Either Recessive or Semi-Dominant X-Linked Retinitis Pigmentosa. Am J Med Genet A 143A:1150-1158.

Zeitz C, Kloeckener-Gruissem B, Forster U, Kohl S, **Magyar I**, Wissinger B, Mátyás G, Borruat FX, Schorderet DF, Zrenner E, Munier FL, Berger W. (2006) Mutations in *CABP4*, the Gene

Encoding the Ca^{2+}-Binding Protein 4, Cause Autosomal Recessive Night Blindness. Am J Hum Genet 79:657-667.

POSTERS

Zweifel SA, Tschuor P, **Magyar I**, Berger W, Becker M, Springer C, Michels S. Multimodal Imaging of Drusen Associated with Malattia Leventinese. Annual Meeting of the Association for Research in Vision and Ophthalmology (ARVO), Fort Lauderdale, Florida, USA. May 1-5, 2011.

Michels S, Zweifel SA, Becker M, Berger W, Tschuor P, **Magyar I**, Springer C. Multimodal Imaging of Drusen Associated with Malattia Leventinese. 34th Annual Macula Society Meeting, Boca Raton, Florida. USA. March 9-12, 2011.

Magyar I, Colman D, Arnold E, Baumgartner D, Bottani A, Fokstuen S, Addor M-C, Berger W, Carrel T, Steinmann B, Mátyás G. Quantitative Sequence Analysis of *FBN1* Premature Termination Codons Provides Evidence for Incomplete NMD in Leukocytes. 8th International Research Symposium on Marfan Syndrome, Warrenton, Virginia, USA, September 11-14, 2010.

Magyar I, Colman D, Arnold E, Baumgartner D, Bottani A, Fokstuen S, Addor M-C, Berger W, Carrel T, Steinmann B, Mátyás G. Quantitative Sequence Analysis of *FBN1* Premature Termination Codons Provides Evidence for Incomplete NMD in Leukocytes. European Human Genetics Conference (ESHG), Vienna, May 23-26, 2009.

Magyar I, Colman D, Arnold E, Baumgartner D, Bottani A, Fokstuen S, Addor M-C, Berger W, Carrel T, Steinmann B, Mátyás G. Quantitative Sequence Analysis Reveals Incomplete Nonsense-Mediated Decay (NMD) of Mutant *FBN1* Transcripts in Leukocytes. 20. Jahrestagung der Deutschen Gesellschaft für Humangenetik (GfHG), Aachen, April 1-3, 2009.

Magyar I, Arnold E, Colman D, Henggeler C, Carrel T, Berger W, Steinmann B, Mátyás G. The Importance of Transcript Analyses: Identification and Characterization of Exonic and Intronic *FBN1* Mutations. 7th Day of clinical research, Zurich, March 27, 2008.

Mátyás G, Alonso S, Patrignani A, Marti M, Arnold E, **Magyar I**, Henggeler C, Carrel T, Steinmann B, Berger W. Large Genomic *FBN1* Deletions Detected by MPLA and SNP Arrays Provide Evidence for True Haploinsufficiency in Marfan Syndrome. ASHG 57th Annual Meeting, San Diego, California, October 23-27, 2007.

Patrignani A, Alonso S, Neuenschwander S, Marti M, Arnold E, **Magyar I**, Henggeler C, Carrel T, Steinmann B, Berger W, Mátyás G. Large Genomic *FBN1* Deletions Detected by MPLA and SNP Arrays Provide Evidence for True Haploinsufficiency in Marfan Syndrome. Inserm workshop 179 «Analysis of CGH and SNP array data», La-Londe-Les-Maures, France, September 20-21, 2007.

Mátyás G, Alonso S, Patrignani A, Marti M, Arnold E, **Magyar I**, Henggeler C, Carrel T, Steinmann B, Berger W. Large Genomic *FBN1* Deletions Detected by MPLA and SNP Arrays Provide Evidence for True Haploinsufficiency in Marfan Syndrome. 6th Day of clinical research, Zurich, March 1-2, 2007.